


Biometric identity systems in law enforcement and the politics of (voice) recognition: The case of SiiP

Big Data & Society
 July–December: 1–13
 © The Author(s) 2021
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/20539517211063604
journals.sagepub.com/home/bds


Fieke Jansen¹ , Javier Sánchez-Monedero² and Lina Dencik¹ 

Abstract

Biometric identity systems are now a prominent feature of contemporary law enforcement, including in Europe. Often advanced on the premise of efficiency and accuracy, they have also been the subject of significant controversy. Much attention has focussed on longer-standing biometric data collection, such as finger-printing and facial recognition, foregrounding concerns with the impact such technologies can have on the nature of policing and fundamental human rights. Less researched is the growing use of voice recognition in law enforcement. This paper examines the case of the recent Speaker Identification Integrated Project, a European wide initiative to create the first international and interoperable database of voice biometrics, now the third largest biometric database at Interpol. Drawing on Freedom of Information requests, interviews and public documentation, we outline the emergence and features of SiiP and explore how voice is recognised and attributed meaning. We understand Speaker Identification Integrated Project as constituting a particular ‘regime of recognition’ premised on the use of soft biometrics (age, language, accent and gender) to disembody voice in order to optimise for difference. This, in turn, has implications for the nature and scope of law enforcement, people’s position in society, and justice concerns more broadly.

Keywords

Speaker identification, biometrics, law enforcement, identity, politics of recognition

Introduction

Biometric technologies are rapidly becoming a central feature of contemporary governance. It follows a trajectory of identity systems such as the introduction of last name, citizen identification number, registration for the purpose of taxation and urban planning, integration of biometric data on passports, and the use of fingerprints and DNA for police investigations that have rendered people legible to a range of corporate and state actors (Scott, 1998; van der Ploeg, 1999). Increasingly, these systems are being complemented by newer forms of biometric technologies that categorize, attribute and establish single identities using digital representations of unique body parts ‘in the wild’ and in ‘real-time’. Facial recognition systems, for example, are now widespread in both private and public spaces, and used in a range of different contexts, from employment to border control to policing (Edri, 2020). Less well known are voice recognition systems, a growing technique used by law enforcement for identifying suspected criminals and terrorists. Within Europe, the Speaker Identification Integrated Project (SiiP)¹ is a prime

example of this development. Launched in 2014 with the aim to build the first international and interoperable database of voice biometrics, it is now the third largest biometric database at Interpol after fingerprints and face (Kofman, 2018). The advent of these biometric identity systems, especially as used by law enforcement, have been the subject of much controversy. Research on facial recognition systems, for example, has documented them as error prone and performing less well on certain demographics, especially black women (Buolamwini and Gebru, 2018) and ‘gender minorities, young and old people, members of the disabled community and manual labourers’ (Kak, 2020). When such systems are used by law enforcement, implications of this disparate impact include identity checks on those

¹Cardiff University, Cardiff, UK

²Córdoba University, Spain

Corresponding author:

Fieke Jansen, Cardiff University, Two Central Square, Central Square, Cardiff, CF10 1FS, UK.

Email: jansenf@cardiff.ac.uk

communities who are already overpoliced (Big Brother Watch, 2018). It further raises questions of fundamental rights such as the right to privacy, the right to freedom of association and assembly, and the right not to be discriminated against. The regulatory frameworks in place to govern these technologies within Europe, such as the European General Data Protection Regulation and the Law Enforcement Directive, have been shown to contain loopholes for the collection of biometrics data of European citizens through identity systems as they rely on broadly defined exceptions for its processing (Kindt, 2020). Several campaigns, particularly in the United States and the United Kingdom, have called for a moratorium on their use, or an outright ban on their presence in public spaces (Kind, 2019).

These controversies point to the role these technologies now play in shaping life chances and opportunities for individuals and different groups in society. They suggest that as *recognition* technologies, the processing of large amounts of biometric data is increasingly part of defining what it means to be recognised in society. Yet how is such recognition understood and what are the terms by which it is dispensed? According to Amoore (2020), algorithmic systems are ‘ethicopolitical’ arrangements that need to be scrutinised for the terms by which they are transforming who and what is made visible and calculable. More specifically, it is important to understand how biometric technologies are creating systems of verification and identification by enabling searchable connections of seemingly disparate objects - a Youtube video, a phone conversation and a police database - to generate new conditions of what someone could be in the world. In this sense, uses of biometric technologies invoke a particular *politics of recognition* that has consequences for individuals’ position in society.

In this paper, we explore the terms of recognition in biometric identity systems by focussing on the novel technical characteristics of the voice identification system SiiP. Still to be fully embedded within law enforcement, our focus is on the research and development phase of SiiP. We explore this as the role of voice in biometric identity systems remains significantly under-researched yet is growing rapidly within law enforcement. Moreover, debates on uses of biometric data by law enforcement have been overwhelmingly speculative, lacking insights about the actual nature and practices of data-based surveillance (Brayne and Christin, 2021). We also consider the focus on voice as particularly pertinent for discussions on a politics of recognition in the context of biometric identity systems as voice carries deep associations with the humanising of people and the capacity of people to give an account of themselves and of their place in the world (Couldry, 2010). In this sense, whilst our understanding of the politics of recognition in relation SiiP builds on broader discussions on biometric identity systems, we are

also confronted with some particularities regarding the role and nature of voice. Drawing on a combination of Freedom of Information requests, interviews with practitioners, white papers, research articles and public documents relating to SiiP, we examine how voice is translated into a recognition technology that attributes and denies forms of recognition. Such processes, we argue, illuminate the power dynamics at play in the advancement of biometric technologies in law enforcement and highlight their pertinence for contemporary struggles for ‘data justice’ (Dencik et al., 2016; Taylor, 2017).

The paper starts by outlining the advent of voice in identity systems, focussing on its use in law enforcement, and the wider geopolitics that contextualise its advancement, before situating these developments in relation to a politics of recognition. Here we are particularly concerned with biometric technologies as ‘regimes of recognition’ (Amoore, 2020) that institutionalise classification systems that in turn shape an agent’s social, cultural and economic position in society. To illustrate the terms of such recognition we then go on to provide an analysis of SiiP, outlining how it attributes characteristics such as age, language, accent and gender to a voice print² upon which inferences are made for the purposes of person identification and classification. Referred to as ‘soft biometrics’ (Abdelwhab and Viriri, 2018; Dantcheva et al., 2015; Kak, 2020), the inference of such attributes as a way to govern populations, we argue, shifts the terms of recognition in such a way that not only targets particular populations, but risks displacing the embodied voice as human account. The growing use of voice identity systems in law enforcement, therefore, invokes new struggles over recognition that need to inform current data justice concerns.

Voice recognition in law enforcement

The expanding biometric industry for military and civilian purposes, including visible facial recognition pilots on the streets, stations and airports across Europe indicate that the use of bodily characteristics such as face, irises or retina for governing are becoming prevalent (Fussey and Murray, 2019; Marciano, 2019). In Scott’s (1998) analysis of the function of the state, identity as a governance instrument is a central component of the emergence of early modern statecraft. It makes society legible to a central authority, allowing the early European nation states to take more informed actions for the purposes of areas such as taxation and security. Since these early ledgers, identity systems have evolved and become more prominent and sophisticated to keep up with the increased global mobility of both goods and people together with the securitisation of politics following the end of the Cold War. Identity systems are therefore important governance tools for monitoring both who enters and exits sovereign territories and how individuals behave domestically (Gates, 2011). Lyon

(2008) refers to the growing use of such systems as ‘governing by identity’, whereby legitimate status in society, access to basic services and public and private space is increasingly tied to the ability to produce and verify someone’s identity. In this sense, identity systems are perceived as a means through which states can more effectively and efficiently engage in statecraft, border control, policing, and administration of public services (Lyon, 2008; van Zoonen, 2013). The more recent turn to biometrics, as a form of governing through identity, originates from the desire of the state to reliably tie a single stable identity to a person (Leese, 2020). The premise is that the use of a digital representation of a finger-, face- or voice print allows for the construction of a reliable single identity that cannot change over time and is less susceptible to exploitation and abuse than other forms of identification (Gates, 2011: 14). Identity systems are therefore increasingly transforming from a unique identification number or a first and last name to being organized around bodily characteristics (Kak, 2020; van Zoonen, 2013).

For law enforcement, identity systems have historically been central to police investigations, both in terms of identifying traces at a crime scene, offenders and victims of crime, as well as in terms of the normative constructions of who or what constitutes a criminal - connecting poor, migrant, and black and brown communities to specific crime priorities. This intersection between race, class and crime is foregrounded in research by Williams and Clarke (2016) who argue that law enforcement invokes identity systems to actively exercise the power to attribute categories of ‘criminality’, or to criminalize specific individuals, communities and groups. The way this attribution of criminality manifests itself in new systems of verification and identification is therefore imperative. With the introduction of more novel biometric technologies in law enforcement we are witnessing a transformation of purpose in the use of identity systems, from being primarily used for investigations to being used for intelligence gathering and policing in real time. In the United Kingdom, for example, police are trialling the use of live facial recognition to monitor shopping streets, events and protests (Big Brother Watch, 2018; Liberty, 2020) and in the United States, half the population is believed to be on facial recognition databases held by law enforcement (Valentino-DeVries, 2020). The prevalence of such applications indicates the extent to which identification is moving from investigation towards the monitoring of specific populations in particular environments.

Whilst the largest known biometric databases are based on fingerprints, and then face, generating considerable attention, there is less research on the use of voice as a biometric to build large-scale systems that can identify and search for speakers. Developments in speaker identification technologies date back to the 1950s and were aimed at supporting security agencies and forensic investigation, and

consisted of a combination of electronic devices and trained experts that were able to analyse the output of these devices to reveal the identity of a person (Pollack et al., 1954). In the 1990s the first automatic voice systems for person identification were created, now used for user-authentication identification and access control key, text-to-speech and speech-to-text assistance (Aronowitz et al., 2011; Rashid et al., 2008). More recently, the increased market interest in voice as a biometrics technology to interface with devices and online services has opened up new possibilities to collect data on how we speak and sound for the purpose of speech, emotion and sentiment recognition and identification (Turow, 2021). These increased voice recognition capabilities are now increasingly commonplace in law enforcement. A survey conducted by Interpol in 2016 found that 44 law enforcement agencies around the world, half of them in Europe, reported they had speaker identification capabilities and 28 respondents answered that they had automatic or human-supervised identification of speakers through audio databases (Morrison et al., 2016).

Recognition and biometric identity systems

Importantly, the construction of identity through many of these biometric-driven recognition technologies is markedly different to how we might otherwise consider identity. Burke and Stets (2009: 3) explain identity as ‘the set of meanings that define who one is when one is an occupant of a particular role in society, a member of a particular group, or claims particular characteristics that identify him or her as a unique person.’ Identities are not static, but are continuously constructed by both individuals to fit the multiple roles they inhabit in society as well as by others in determining the role a person occupies in society. The ability to construct identities is often regarded as central for people to relate to each other through community and group memberships and to understand the world around them (Young, 2011). With the advent of data-intensive technologies, identity is said to be increasingly ‘algorithmic’ (Cheney-Lippold, 2017) and often digitally mediated, such as through social media platforms and apps, and constructed through data-driven forms of profiling that create ‘data doubles’ (Haggerty and Ericson, 2000). This construction of algorithmic identities, that enable a multiplicity of categorization and social sorting, often happens outside the consciousness and control of individuals, communities and groups (Andrejevic, 2012).

In biometric recognition systems, the construction of identities relies on the attribution of so-called ‘soft biometrics’ that refers to the inference of ‘demographic characteristics, emotional states and personality traits from bodily data’ (Kak, 2020: 6). Here the aim is not only to tie a

single stable identity to a digital representation of a fingerprint, face and voice for the purpose of registration, authentication and identification, but also to infer certain characteristics from a biometric trait to inform classification. This, in turn, constitutes actionable intelligence. As Gates (2011: 15) argues, biometric systems not only standardise identity systems but ‘push those standardised categories of identities back out onto the individual’ for a specific purpose. Amoore (2020) therefore sees recognition technologies as embodying the determination of what is ‘useful’ or who is recognizable as a target of interest. That is, she argues, machine learning algorithms embody a ‘regime of recognition’ that identifies who or what matters to the event. Drawing on Fanon’s concept of ‘epidermalization’, Browne (2015) refers to these novel biometric systems as ‘digital epidermalization’ that serve to ascribe meanings to certain bodies from a disembodied gaze. More specifically, rooted in ‘prototypical whiteness’ biometric technologies are inscribed in racialising schemas that make some bodies problematic and not others.

Identity systems based on representations of voice are therefore part of a suite of biometric technologies that set out particular terms of recognizability of a body as human, as ‘fully political’ (Amoore, 2020: 4). They are not merely validating or attributing truth to an existing identity, but actively produce and institutionalise classification systems that exert attributions of who can be recognised and for what. When deployed in the context of law enforcement, such biometric recognition technologies extend these attributions onto individuals with potentially significant impact on their position in society. In this sense, they are part of the mechanisms through which categorisation is established and claims to the social world are made, constituting an important aspect of what Bourdieu (1982/2018) refers to as ‘struggles over classification’. Such struggles necessitate questions of who and how we classify, and what classifications have the most social weight. These questions, in turn, are key for understanding the conditions within which individuals form and engage with a ‘politics of recognition’ (Taylor, 1994).

Here we do not intend to carry out a lengthy discussion on the many different interpretations of (the politics of) recognition as advanced by different scholars (see McBride, 2013 for a useful overview), but refer to it as a way to highlight the significance of uses of biometric technologies in law enforcement for how individuals come to understand themselves in relation to others and the material and symbolic inequalities that affect the formation of such recognition (Herzig, 2017). In particular, following Fourcade and Healy (2017), we understand the algorithmic processing of data as scoring and ranking instruments that contribute to the accumulation of symbolic and intangible capital that in turn shape an agent’s social, cultural and economic position in society. An individual’s value or worth, in this sense, is construed through the meaning they

afford algorithmic classification systems. Moreover, drawing on McBride (2013), the politics of recognition in this context is not so much a question of interpersonal relations as one of the social power needed to gain authority to dispense recognition. That is, the meaning attributed to recognition through biometric identity systems is compounded by existing social relations that condition how and by whom recognition is distributed in society. This is especially pertinent for algorithmic systems deployed for the purposes of policing and law enforcement.

Here the focus on voice may be particularly relevant for thinking about recognition in that it extends existing discussions on biometrics, but also introduces certain features and ambivalences that are particular to voice. Even if not always made explicit, the relationship between recognition and voice is intimately linked in understandings of justice, in that voice is often assumed to be one of the main mechanisms through which people are able to give an account of themselves and their place in the world (Couldry, 2010). At the same time, voice is often seen to be subject to manipulation and silencing in ways that make its abstraction and use particularly important. This may especially be the case where voice is disembedded from embodied contexts as in the case of biometric recognition technologies. In what follows, we therefore engage with the way terms of recognition might be transformed through the growing use of biometric identification systems by focussing particularly on the practices and premises imbued in SiiP, the first voice biometric database with international scope and a project that offers new technical novelties based on Open-Source intelligence (OSINT) and the use of ‘soft biometrics’ to classify voice samples. In doing so, we are concerned with what Lyon (2008) describes as ‘governing by identity’ that invokes an ethics of classification requiring political scrutiny (see also van Zoonen, 2013). As we go on to argue, we see such scrutiny as especially relevant for the displacement and appropriation of voice enacted through biometric identity systems as part of contemporary struggles over recognition.

The case of SiiP

Whilst voice as biometric data to verify a person has been used for decades, unknown speaker recognition is a more nascent field that has been made possible through technological advances and the rapid proliferation of audio samples that can be extracted from video platforms, social media platforms and mobile phones. SiiP is a pertinent case for studying this development as it has become prevalent amongst law enforcement agencies internationally as a way to overcome the increased use of multiple and arbitrary identities by suspected criminals and terrorists. As a project SiiP itself has completed and has been integrated into Interpol’s biometric databases with further incarnations of voice recognition systems being developed, such as

Roxanne,³ a next generation research project on biometrics in law enforcement that aims to combine voice recognition, language and video technologies for the purposes of, amongst other things, social graph analysis. This buildout indicates the significance voice recognition technologies are having, beyond the focus on SiiP as outlined here.

Doing research on developments in policing is notoriously difficult in terms of access (Brayne and Christin, 2021), so in order to explore the case of SiiP we draw on a combination of methods that provide insights into the nature and uses of the system. These include five freedom of information requests sent to the European Research Executive Agency (REA), the London Metropolitan police, the German Bundeskriminalamt, the Italian Ministero della Difesa and Portuguese Ministério da Justiça, two interviews with law enforcement practitioners directly involved in the SiiP project, and publicly available project documentation and research papers to outline the development and key features of SiiP. The FOIs requested access to the REA ethics report, evaluation documents, and deliverable documents,⁴ and national law enforcement partner documents concerning the piloting of SiiP, specific data sources used and the plans for future implementation. The latter was declined on national security grounds, however some referred to publicly available information on gatherings organized by Interpol. This was complemented by research papers on the technical construction of identity based on soft biometrics (Ferras et al., 2016; Khelif et al., 2017, 2018) and public documentation available about the project from the European Commission.⁵ Information on further details of the whole system, use cases, privacy by design considerations and review of ethical and societal aspects were extracted from project reports available in the archived copy of the website. In analysing this data, we are particularly seeking to understand voice as a praxis of biometric identification in the context of law enforcement and how this is shaping the scope and remit of crime. This in turn informs our exploration of the terms upon which voice is 'recognised' within the identification system and the meanings attached to such recognition. We therefore start by outlining the context out of which SiiP has emerged in law enforcement and the rationale provided for its uses, before turning to some of the key features and premises of SiiP that help to illustrate how recognition is operationalised through biometric technologies.

The emergence of SiiP

From the outset SiiP has been positioned as a novel biometric identity system for law enforcement developed to address a pressing organizational need and to increase the ability to respond to the security priorities of organized crime and terrorism. As a research project, SiiP was funded by the European Union under its FP7-Security

programme with the aim to build the first international and interoperable database of voice biometrics to support investigations into transnational threats, terrorism, and organized crime. Verint Systems Ltd, an Israeli military and security company, coordinated the project, Nuance Communication (formerly known as Loquenco), SAIL LABS and Idiap provided the technical capacity, and Interpol, the Bundeskriminalamt, London Metropolitan Police, Ministero della Difesa and Ministério da Justiça were the participating law enforcement agencies. In its funding proposal to the European Commission this consortium claim that one of the most prominent contemporary obstacles to the ability of law enforcement to fight against organized crime and terrorism is the fact that terrorists and criminals can use multiple and arbitrary identities. The use of encrypted communication and burner phones allows criminals and terrorists to hide their real identity, and social media and communication apps allow them to easily take on many different identities and nicknames, all of which makes it increasingly more difficult for law enforcement to track or monitor perceived criminal activity. SiiP is said to facilitate the identification of criminals and suspected terrorists as well as map and trace future networks of suspects. As such, the emergence of SiiP needs to be understood both in terms of the defined security priorities of the European Union which are reflected in specific FP7 Security programmes, as well as the continuous development of biometric identity systems within the security industry that aligns with organizational challenges.

The SiiP project aims to develop 'a system that identifies voices in audio sourced from lawfully intercepted communication and social media' (Interpol, 2018b) and is said to assist law enforcement in a number of use cases ranging from identifying terrorism suspects to perpetrators of child pornography. The consortium behind the project argues that in the case where terror suspects have covered their faces while committing a violent act but were caught on camera, the voice samples of the unknown suspects can be cross-referenced against a database of known terror suspects. In the case where an unknown masked person posted a video of themselves committing a terror act a system like SiiP would allow law enforcement to compare this unknown voice reference collected from social media against other publicly available social media videos. If videos with matching voice references are found that include unmasked images of the suspect, this would in turn facilitate identification by law enforcement. Another use case put forward is in relation to organized crime and the use of multiple online and offline identities. According to the documentation, SiiP would allow law enforcement to identify a speaker in phone tabs or identify someone who is trying to enter a country under a false identity. When a border officer suspects the use of a false identity, a voice sample from an interview can be compared in real time to those voice samples of known criminals that are

stored at the police headquarters or to the voice samples stored at Interpol. A third use case is the use of voice samples to identify the perpetrator or victims of child sexual exploitation by matching the voice of unknown perpetrators in the background of confiscated video evidence against a larger database.

In all these use cases it is argued in the project documentation that ‘SiiP could make voice identification a more reliable and powerful part of a criminal investigation’ (Interpol, 2018b). As such, the novel features of SiiP are explicitly linked to crime investigation uses rather than other types of uses (such as intelligence). This is important as presenting SiiP within a contained field of crime investigation may bypass engagement with key questions and concerns that have marked debates about the uses of biometric technologies in law enforcement otherwise. However, as noted above and as we go on to argue, the reliance on soft biometrics as a way to narrow down the search field actively undermines any clear distinction between investigation and intelligence. Moreover, these use cases foreground representations of voice as a versatile basis for identification; not only can it be applied to a range of predetermined crime priorities it can also be used to match an unknown voice to a diffuse corpus of data, ranging from known voice prints within law enforcement databases, to unknown voice samples in telecommunications infrastructure, collected from social media (OSINT) or in physical demarcated spaces (borders). This uniquely positions voice in terms of what can be digitally captured and analysed as the basis for recognition.

Features of SiiP

In our interviews, the practitioners highlighted a number of novel features in the SiiP system, the most notable being the use of soft biometrics to increase the accuracy of voice recognition. SiiP offers the use of 7 different ‘engines’ to identify a suspect: voiceprints recognition, keyword spotting, voice cloning detection, and age, language, accent and gender identification (European Commission, 2017). This is a notable upgrade from previously used voice identification systems which only had the capability of combining two engines, such as age and language or accents. The move to expand and combine a multiplicity of soft biometric characteristics is seen as one way to establish more reliability in the identification of a suspected criminal or terrorist. Furthermore, it is suggested in the project description that SiiP would be capable of filtering out voice samples of family members or individuals that use the suspects’ phone and thereby decrease surveillance of ‘innocents’ (European Commission, 2017). This suggests that increased accuracy through the attribution of soft biometric will allow law enforcement to make some voices visible and others invisible. Importantly, the aim is not to determine a singular identity but to offer multiple (3–5) matches that

allow for a range of options. This, it was claimed in interviews by practitioners, ensures humans remain ‘in-the-loop’ for any decision-making, taking account of the uncertainties surrounding voice identification. Unlike fingerprints or DNA, the voice characteristics of a person will differ depending on time of day, levels of stress and other externalities (e.g. smoking).

Practitioners therefore rely on soft biometrics - the attribution of characteristics such as age, language, accent and gender to a voice print upon which inferences are made - both for the analysis of an unknown voice as well as a pre-selection mechanism to narrow the dataset collected from OSINT (predominantly social media) for a 1:N comparison. This is explained in project documentation from Interpol as ‘in open-source intelligence you can use all these machines to for example narrow down space where you are looking for a speaker. Say you are looking for a speaker that speaks Arabic, is male and is an adult, but he has an accent [...] like Saudi Arabia, you could narrow down the search space into these features and hopefully you can narrow it down and maybe the chance is better to really find this fish in the sea’ (Interpol, 2018a). The application of soft biometrics for the purpose of pre-selection is a novel feature in SiiP that serves to create group classifications of potential ‘criminal’ or ‘terrorist’ based on extracted samples of digital representations of voice that share personal characteristics with a pre-defined profile.

Speaker verification and identification systems generally consist of a pipeline of several steps including feature extraction, speaker modelling, scoring and, in the case of verification, decision-making. First a system performs feature extraction to convert the speech signal of an audio utterance to a set of feature vectors that can be processed by a computer. Feature vectors are processed to create text and channel independent representations of a person known as ‘speaker models’. A reliable system should generate very similar speaker models from different audio samples of the same person so, typically, there would be several vectors for each person to represent inter-speaker variability (Dehak et al., 2011). State-of-the-art software, such as the one in SiiP, uses a technique called i-vectors (identity vectors) as speaker models together with statistical and machine learning techniques to compare and score the similarity of these models (Kheif et al., 2017). Soft biometrics models are codified in a similar way through Universal Background Models, which are speaker independent models that can be compared against a person-specific model (Li and Jain, 2015).

Although SiiP explored different statistical and machine learning techniques for each step of the pipeline, we can point out some general assumptions and logics of these techniques when building computational representations of identities and soft biometrics. To improve accuracy, these techniques follow two general principles of machine learning: the computational representation of items of the

same class have to be similar whilst the representation of samples of different classes have to be very dissimilar to other groups. To seek for these principles, statistical techniques such as factor analysis or linear discriminant analysis are used to find a vectorized representation in which the ratio between between-class variance (different demographic groups, e.g. samples of adults vs. children) and within-class variance (same demographic group, e.g. samples of children) is maximized (Li and Jain, 2015). This generic principle for data transformation is also present in the machine learning methods used to build the final scoring tools such as support vector machines and deep neural networks. This means that audio characteristics that are similar amongst demographic groups will be removed from the model whilst characteristics which better differentiate between categories will have a stronger influence in the score that measures how likely a person belongs to a demographic group. The above reasoning is valid to understand the optimization logic of the statistical techniques however this does not necessarily mean that the resulting system will meet these properties for all the identity classification subsystems. Actual systems working with real data can score persons within a gradient of values that overlap between categories rather than producing clear binary groups.

Previous systems relied on information fusion of different speaker recognition engines to improve accuracy, what is described as intra-task fusion, while SiiP exploits heterogeneous systems for gender, age, language and accent identification extended by speech-to-text transcription with keyword spotting, so-called inter-task fusion. The overview of identity information fusion as used in SiiP is shown in Figure 1.

The use of soft biometrics classification models has previously been developed in the field of computer vision and facial recognition to improve recognition rates through inter-task fusion or to describe persons in video and images (see for example Park and Jain, 2010 on visual surveillance and forensics). The demographic and keyword information used in SiiP will generally be available as metadata associated with a speaker in the context of an investigation. The underlying logic is that when comparing two speaker models they should be considered closer in the model space if the inferred characteristics are similar. For example, a person identified as an adult man with a Hindi accent would more likely match audio samples with these inferred characteristics. The developers of SiiP examined several ways to fuse speaker information with accent, language and gender (Ferrás et al., 2016) and also explored statistical techniques to infer demographic characteristics when there is uncertainty about or absence of demographic information in metadata (Madikeri et al., 2019). According to the authors, the experimental study concluded that adding this side information reduced the error rate to 0.5% on benchmark datasets (Ferrás et al., 2016). The

available documentation indicates that the SiiP language identification can recognise 22 languages and identify several English accents (Native, Chinese, Russian, Hindi and Korean), while practitioners highlighted that next to English they also considered other languages such as Turkish, Portuguese and Arabic. These modules are trained to discriminate between languages and accents and avoid identifying acoustic features from speakers. There is one module for age identification and gender identification that categorizes each speaker with a binary label, where age is reduced to ‘child’ or ‘adult’ and gender to ‘men’ or ‘women’. Both modules are implemented with machine learning models trained on a combination of English and German corpora. The SiiP project did not create a new dataset but relied on previous ones associated with other speaker identification systems and previous academic work. As such, whilst using a multi-source dataset can strengthen the robustness of the system, this also means that the categories associated with a person’s identity are driven by data availability, and the terms of data collection, above other criteria.

The second novel aspect of SiiP pointed out in our interviews with practitioners is to simplify law enforcement use of a variety of data sources, such as the integration of data collection from OSINT, capture voice data from audio recording from mobile and satellite, and the use of mobile phones to record voice samples and cross reference them to databases at law enforcement headquarters and Interpol. Features of Nuance’s products allow for audio mining, which entails grouping similar speaker voices in a database to associate unknown data in large sets of information.⁶ To deal with sound variability, the development team created a ‘Lawful Interception Simulator’ to test the system performance when dealing with different communication channels (SATCOM, PSTN, cellular, telecom VOIP and Internet VOIP apps). An investigator can query these systems through the SiiP Portal to find voice samples beyond basic keyword searches by including language, geo-location, taxonomy, entity associations etc. and the search results can be incorporated to the SiiP database including not only audio information, with associated metadata, but also other captured multimedia information (video, photographs and images) that will be maintained for possible evidential purposes.

Importantly, as noted above, when relying on OSINT or the interception of phone calls for the collection of voice samples, which are captured ‘in the wild’ for multiple unknown speakers, it is necessary to account for environmental factors to trust an accuracy rate. However, the available documentation on the SiiP project does not indicate the actual performance of the system with real data and what characteristics of an audio sample such as length or quality are enough to identify a person in a large OSINT database or phone recordings. According to research, a small audio sample of 30–60 s length can be enough to

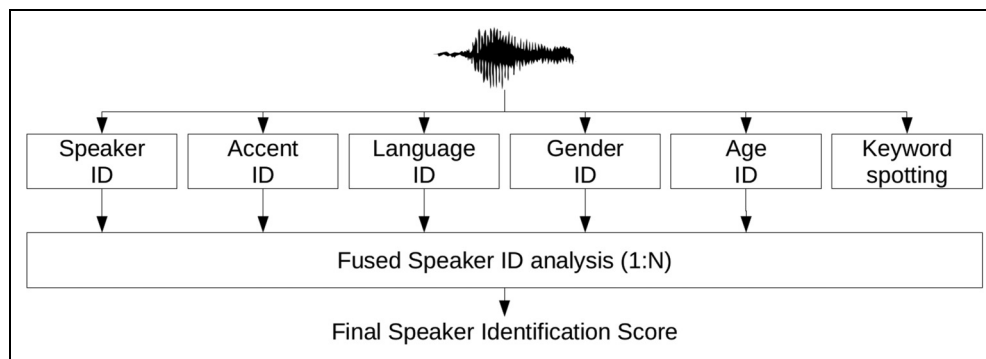


Figure 1. Diagram showing the information fusion procedure of SiiP. Adapted from Khelif et al. (2018).

verify the identity of a person in benchmark datasets (Poddar et al., 2019) yet the robustness of the tools depends on factors such as noise, heterogeneous speakers, heterogeneous recording devices or audio encoding.⁷ Practitioners recognised the quality of voice samples needed for speaker identification as one of the key challenges with the project, OSINT data generating better results than phone recordings. Moreover, in the context of security and law enforcement agencies a speaker can adopt counter measures. In biometrics, spoofing or presentation attacks consist of altering the data before it is captured/measured to evade speaker identification (Alegre et al., 2014) or engage in mimicry of the identity of a person (Li et al., 2020).

SiiP, therefore, advances key assumptions about not only the nature and uses of biometric data by privileging the abstraction of demographic characteristics (soft biometrics) as a way to infer and classify identities, but also the relationship between digital representations of voice and legal categorisations of action distinguishing suspect and non-suspect by relying on OSINT as the primary source for data collection to inform actionable intelligence. The case of SiiP is therefore a pertinent example of the way biometric technologies are increasingly integrated into social practices around some significant operational logics that in turn inform how and on what terms a person is recognised. As we go on to discuss below, SiiP can in this regard be seen as illustrative of a particular politics of recognition that emerges from the growing reliance on biometric identity systems based on voice.

The politics of (voice) recognition

The emergence and features of SiiP is part of a broader trend that elevates the role of biometric technologies in shaping life opportunities. This is evident from the active construction of identities through soft biometrics abstracted from a vast range of data sources that inform classifications of political categories such as innocent/criminal/terrorist.

As Browne (2015) points out, the use of biometric data follows a genealogy of ‘branding’ (rooted in slavery) in which the application of biometrics is in the verification, identification, and automation practices that enable the body to function as evidence. Furthermore, these digitised instances need to be understood in the context of specific historical and institutional sites within specific discursive formations and practices, by specific enunciative strategies. In analysing SiiP, we can see that while debates have primarily centred around the use of facial recognition, a new form of biometric technology has emerged in the context of law enforcement that further extends the body as evidence and contributes to increased levels of real time monitoring. As a project, SiiP seeks to not merely identify voices, but rather to ‘actively generate recognizability’ (Amoore, 2020: 69) in terms of who or what can be registered as of interest through an embedded authority to dispense or relinquish recognition; an authority that does not reside wholly in a recognizable human. Beyond the use cases provided in the project documentation, we still know little about how SiiP is used by law enforcement agencies in practice, but the integration of SiiP within Interpol and the advancement of voice-based data systems in other security research projects (such as Roxanne) indicates a growing significance of generating recognisability through voice. This suggests that law enforcement agencies have an interest in the affordances of SiiP to construct recognizable identities through the automated processing of digitised data that sets the parameters for the boundaries of crime, actionable intelligence and subsequent decision-making. What matters in terms of recognizability are the digital representations of voice over and above any embodied voice in any particular space. This shifts the terms of recognition and the way in which struggles over recognition might emerge.

Debates on the political implications of biometric identity systems such as finger-printing and digital imaging based on face or gait have highlighted concerns about surveillance as well as the high error-rates and discriminatory

outputs of such technologies. Emphasising the ‘prototypical whiteness’ (Browne, 2015) that has marked the genealogy of biometric technologies, these concerns have pointed to the rate of misidentification of ‘othered’ faces, such as black and brown, women and queer communities (Brayne, 2020, Buolamwini and Gebru, 2018). Such observations are particularly relevant in systems where abstracting biometric data is a condition for the recognition of basic rights, and where misidentification can lead to increased police surveillance and other actions. Importantly, whilst showcasing the extent of ‘biases’ in biometric identity systems has served to undermine the legitimacy of their uses, there has also been a worry that it may spur on technical fixes that seek to optimize facial recognition for those very groups prone to misidentification. We see this worry manifest itself in the case of SiiP as a system of verification and identification that does not optimize to recognize a mere voice but optimizes to recognize the voice of specific communities. Soft biometrics are used as proxies for demographic categories to improve the accuracy of the system from the perspective of law enforcement, attributing voice characteristics to categories of criminality. Training a system to identify specific accents, languages and genders that are associated with the recognition of a ‘terrorist’ in itself makes certain individuals and communities more visible to law enforcement. As such, the integration of voice samples into databases for the purposes of law enforcement embeds additional layers of conditionality and ontological insecurity (a body made out of place [Browne, 2015]) into the regime of recognition constructed by biometric identity systems.

Furthermore, the reductionist categories of binary labels inferred by soft biometrics rely on criteria based on already available data that is unlikely to match the cultural and social diversity of the general population. Such uses of soft biometrics have been shown to lack scientific validation and often rely on deeply contested assumptions about physiological attributes and the relationship between people and data (Sánchez-Monedero and Dencik, 2020). Pre-selecting the digital representations of some voices collected from data infrastructures to inform the labelling of criminals or terrorists invokes a particular regime of recognition, in which institutional priorities shape what dispersed groups of individuals are recognized and algorithmically processed by law enforcement on the basis of sharing certain voice characteristics. Thus this process of optimizing for biometric differences transforms the everyday online expression of some communities into sites for criminal investigation. In turn, the use of soft biometrics as a pre-selection mechanism risks embedding forms of stereotyping in institutions and obscuring the criminalization of specific communities (Williams, 2015). This algorithmic processing of voice happens without knowledge of those impacted, ensuring and entrenching the obscurity of the process through which specific communities are

criminalised. While the SiiP project advances the argument that the use of soft biometrics will increase the accuracy of identifying suspects and also decrease surveillance of ‘innocents’, the pre-selection of certain voices highlights the extent to which such categories are necessarily premised on differential distribution of recognition amongst populations. It compels the question of how voices are recognised as a threat (criminal or terrorist) and how voices are deemed innocent.

What is more, digital representations of voice are well known to be prone to environmental interferences that make them unreliable. Practitioners engaged in the SiiP project try to overcome this by relying more on some type of data rather than other (e.g. OSINT rather than phone tapings), but in order to perform its claimed tasks, the system still necessarily invokes certain ‘ground-truths’ that facilitates a process of actualising a possibility of potentials into one, without making this explicit. Amoore (2019) outlines it as a condition of ‘algorithmic doubt and certainty’ in which algorithmic systems erase the presence of doubt whilst simultaneously generating ‘the parameters against which uncertainty will be adjudicated’ so that a multiplicity of doubts can be condensed to a single output beyond doubt. By ignoring the fallibilities of what the algorithm has learned about the world, by erasing doubt as an inherent feature of knowledge, uncertainties, errors and ‘biases’ in biometric identity constructions based on voice are baked into the system from the outset along with the political consequences that emerge from that.

Although it has been very under-researched, the focus on voice is particularly pertinent in considering biometric identity systems as regimes of recognition. Here, voice as a technology that optimizes for difference allows us to situate it in a broader ‘struggle over classification’ (Bourdieu, 1982/2018), where the political rationale of who and how people and communities get classified as subjects of interest in turn has the potential to shape their social, cultural and economic position in society (Fourcade and Healy, 2017). As the case of SiiP illustrates, the collection of voice samples across data sources (social media, mobile phones etc.) and the integration of voice data into shared databases held by law enforcement agencies actively disembods voice from body and context (across the general population, on a global scale). The use of (real time) voice sample collection allows any place and data infrastructure to become a crime scene, a site for investigation where one can lawfully record a person. Conditions are therefore created where digital representations of some voices can be abstracted, circulated and classified with disregard for the contextualised and embodied meaning of these voices. Social media platforms, for example, shape conditions within which (mis)recognition might be carried out not necessarily as sites of identity formation, but as an intrinsic part of political and institutional crime priorities, disenfranchising and reappropriating certain voices. This

disembedding, or dispossession (Couldry and Mejias, 2019), of voice in turn serves to undermine the value of voice, as in the capacity for people to give an account of themselves and their place in the world. Instead, biometric information increasingly displaces the ‘muted body’ (Browne, 2015) - any claim to an identity’s truth resides overwhelmingly in the inferred characteristics of biometric data that end up speaking for an individual and their position in society.

The politics of (voice) recognition that emerge from the uses of biometric identity systems in law enforcement are therefore significant for the ways in which datafication intersects with justice concerns - or data justice - in several respects. They highlight the need to unpack not only who is recognized by data systems but also how that recognizability is constructed and with what implications for people’s life chances. Voice in both its symbolic and material form is often regarded as fundamental to how people are able to make claims for themselves and the environment they inhabit, yet the terms upon which voice comes to matter is a political question (Couldry, 2010). The advent of biometric identity systems such as SiiP radically transforms the relationship between voice and recognition. Rather than voice being a channel through which recognition can be sought based on an individual’s claims to experiences and the social world, SiiP illustrates a process of alienation between voice and recognition; the embodied voice no longer having any meaningful claims to the terms of recognition and how recognition might be attributed. Moreover, SiiP subverts this relationship through disembedding voice and appropriating spaces of expression of voice into spaces of classification of voice for purposes of law enforcement. In this sense, biometric identity systems can be seen as inherently misrecognising; a form of misrecognition that in the case of SiiP criminalizes individuals who share voice characteristics attributed to a crime priority.

The case of SiiP and the focus on voice recognition in law enforcement therefore extends discussions on data justice beyond an engagement with rights and freedoms, and forces us to contemplate the terms by which we are able to give an account of ourselves and how we come to be recognised in the world. Although presented in terms of ‘research’ as a way to neutralise its role as an instrument of governance, the advancement of SiiP is deeply embedded in the current political and strategic vision within Europe. This vision entails the prioritization of experimentation on specific crime types over others (Garland, 2004; Williams, 2015), the desire of law enforcement to not only utilize new technologies but also expand the spatio-temporal horizon of what constitutes a crime scene and who constitutes a potential suspect through such exploitation of available technical capabilities. The politics that emerges from the development and uses of SiiP is therefore indicative of much broader transformations in the

conditions within which struggles for recognition are pursued today.

Conclusion

The reliance on different kinds of biometric data for investigating crime and other illegal activity has a long and contentious history. With the rise of automated machine learning systems in policing that abstract biometric information from artefacts such as images or sensors, biometric identity systems have become the subject of significant scrutiny and debate. Whilst developers and law enforcement bodies have often advanced such technologies with a promise of more efficient and accurate practices, concerns have been raised about the impact the deployment of biometric identity systems in policing might have on fundamental human rights and the policing of different communities. These have often focussed on the nature and scale of surveillance that the use of such technologies entails and what this might mean for the right to individual privacy, freedom of expression and freedom of assembly. In addition, prominent research into biometric identity systems, especially facial recognition technologies, has shown they are prone to high error-rates, particularly amongst minority groups that are subject to increased risk of criminalization when such technologies are used. This has positioned biometric identity systems as instruments of governance that are important for concerns with justice.

The increasing use of voice samples to inform biometric identity systems deployed by law enforcement provides further ammunition to these debates. Whilst voice recognition technologies have received much less public and scholarly attention, they are rapidly becoming part of this new regime of recognition that determines who is made visible and calculable. The case of SiiP provides some critical insights into the terms on which such recognition is carried out and attributed meaning. Relying on inferences based on soft biometrics, such as gender, language and accents, the SiiP project is advanced on the premise of increasing accuracy and addressing key organisational needs. Accuracy and need here is understood not just in terms of establishing a single identity, but rather to create a classification system based on matching voice characteristics to crime priorities. This classification system is, in turn, conditioned by the availability of data in the form of digital representations of voice across online platforms and mobile phones.

The integration of SiiP in law enforcement is therefore significant for understanding the conditions within which contemporary politics of recognition play out. In abstracting voice from data infrastructures as a way to inform recognisability understood as who or what may be of interest to law enforcement, voice is disembedded from the embodied and social context that gives it meaning as a vehicle for how a person gives an account of themselves

and their place in the world. Instead, the value of voice as it relates to recognition is determined by the calculable inferences that can be made from it in relation to a pre-defined objective (crime priority). This, in turn, serves to optimise for difference so as to better recognise particular voices over others. Voice, in this sense, is recognizable only insofar as it provides value to categories of actionable intelligence (criminal/terrorist/innocent) that in turn impacts on people's position in society. This suggests there is an inherent misrecognition - and therefore injustice - that accompanies the displacement of voice in biometric identity systems. As voice recognition technologies are set to become increasingly significant for law enforcement and governance structures, moving into areas of social network analysis, such scrutiny of emerging recognition regimes and the way they impact on policing practices needs to form a much bigger part of our understanding of data justice.

Acknowledgements

Research for this article has been supported by the ERC Starting Grant DATAJUSTICE (grant no. 759903) under the Horizon 2020 research and innovation programme. We would also like to thank the three anonymous reviewers and Jędrzej Niklas for his comments on an earlier version of this article.

Notes on contributors

Fieke Jansen is a PhD candidate at Cardiff University's School of Journalism, Media and Culture and the Data Justice Lab. She is interested in re-politicizing data and technology, by understanding its historical, social, cultural and political context in Europe. Her research focuses on the impact of implementing data-driven decision-making in European police forces on targeted communities. Prior to starting her PhD, Fieke worked as a practitioner on the intersection of human rights, internet, and artificial intelligence. She can be reached at jansenf@cardiff.ac.uk

Javier Sánchez Monedero is a Distinguished Researcher (Beatriz Galindo grant) at the AYRNA research group in the Dept. of Computer Science at the University of Córdoba and an affiliate researcher of the Data Justice Lab at the University of Cardiff. With a background in computer science, his current research aims to fill the knowledge gap between social and media researchers and technology as well as to perform technological auditing and design proposals in the intersection of intelligent information systems and social justice. Javier has worked on several projects involving distributed systems and machine learning targeting problems of biomedicine, renewable energy and climatology among others. He can be reached at jsanchezm@uco.es.

Lina Dencik is Professor at Cardiff University's School of Journalism, Media and Culture and Co-Director of the Data Justice Lab. Her research concerns the interplay between media developments and social and political change, with a particular focus on governance and resistance. She has published widely in the areas of digital media and the politics of data and is the author of several books, most recently *Digital Citizenship in a*

Datified Society (with Arne Hintz and Karin Wahl-Jorgensen, Polity Press 2018) and *The Media Manifesto* (with Natalie Fenton, Des Freedman and Justin Schlosberg, Polity Press 2020). Lina is Principal Investigator on the ERC-funded DATAJUSTICE project. She can be reached at DencikL@cardiff.ac.uk.

Declaration of conflicting interests


The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the H2020 European Research Council (grant number 759903).

ORCID iDs

Fieke Jansen  <https://orcid.org/0000-0002-1295-7788>

Lina Dencik  <https://orcid.org/0000-0002-1982-0901>

Notes

1. <https://cordis.europa.eu/project/id/607784>
2. Voice print is the term we use to distinguish between a voice sample that is collected from and through data infrastructures and the abstracted voice that has been analysed.
3. <https://cordis.europa.eu/project/id/833635>
4. Deliverable documents in the time period of the testing of the project with the four law enforcement partners; Ministério da Justiça (Portugal), Ministero Della Difesa (Italy), Mayor's Office for Policing and Crime (United Kingdom), and the Bundeskriminalamt (Germany). https://www.asktheeu.org/en/request/information_related_to_the_h2020#outgoing-15558
5. Way Back Machine archived material of SiiP website <https://web.archive.org/web/20181229171923/http://www.siiip.eu/>
6. https://www.nuance.com/content/dam/nuance/en_us/collateral/enterprise/data-sheet/ds-nuance-forensics-en-us.pdf.
7. https://www.asktheeu.org/en/request/information_related_to_the_h2020

References

- Abdelwhab A and Viriri S (2018) A survey on soft biometrics for human identification. In: Yang J, Park DS, Yoon S, Chen Y and Zhang C (eds) *Machine Learning and Biometrics*. London: IntechOpen, pp.37–56.
- Alegre F, Soldi G and Evans N (2014) Evasion and obfuscation in automatic speaker verification. In: ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings, Florence, Italy, 4–9 May 2014, pp.749–753.
- Amoore L (2019) Doubt and the algorithm: On the partial accounts of machine learning. *Theory, Culture & Society* 36(6): 147–169.
- Amoore L (2020) *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Durham: Duke University Press.
- Andrejevic M (2012) Exploitation in the data mine. In: Fuchs C, Boersma K, Albrechtslund A and Sandoval M (eds) *Internet*

- and Surveillance. New York and London: Routledge, pp.91–108.
- Aronowitz H, Hoory R, Pelecanos J, et al. (2011) New developments in voice biometrics for user authentication. In: Twelfth annual conference of the international speech communication association, Florence, Italy, 28–31 August 2011, pp.17–20.
- Big Brother Watch (2018) Face off: The lawless growth of facial recognition in UK policing. Report, Big Brother Watch, UK, May. Available at: <https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf>.
- Bourdieu P (1982/2018) *Classification Struggles*. Cambridge and Medford, MA: Polity.
- Brayne S (2020) *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York: Oxford University Press.
- Brayne S and Christin A (2021) Technologies of crime prediction: The reception of algorithms in policing and criminal courts. *Social Problems* 68(3): 608–624.
- Browne S (2015) *Dark Matters: On the Surveillance of Blackness*. Durham: Duke University Press.
- Buolamwini J and Gebu T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st conference on fairness, accountability and transparency in machine learning research 81, pp.77–91.
- Burke PJ and Stets JE (2009) *Identity Theory*. Oxford: Oxford University Press.
- Cheney-Lippold J (2017) *We Are Data; Algorithms and the Making of Our Digital Selves*. New York: New York University Press.
- Couldry N (2010) *Why Voice Matters: Culture and Politics After Neoliberalism*. London: Sage publications.
- Couldry N and Mejias U (2019) Data colonialism: Rethinking big data's relationship to the colonial subject. *Television and New Media* 20(4): 336–349.
- Dantcheva A, Elia P and Ross A (2015) What else does your biometric data reveal? A survey on soft biometrics. *IEEE Transactions on Information Forensics and Security* 11(3): 441–467.
- Dehak N, Kenny PJ, Dehak R, et al. (2011) Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing* 19(4): 788–798.
- Dencik L, Hintz A and Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society* 3(2): 1–12.
- Edri (2020) Facial recognition & biometric surveillance: document pool. Available at: <https://edri.org/our-work/facial-recognition-document-pool/> (accessed 4 June 2021).
- European Commission (2017) Speaker identification integrated project. September 10th 2017. Available at: <https://cordis.europa.eu/project/id/607784> (accessed 17 June 2021).
- Ferras M, Madikeri SR, Dey S, et al. (2016) Inter-task system fusion for speaker recognition. In: Interspeech 2016, San Francisco, USA, 8–12 September, pp.1810–1814.
- Fourcade M and Healy K (2017) Seeing like a market. *Socio-Economic Review* 15(1): 9–29.
- Fussey P and Murray D (2019). Independent report on the London Metropolitan Police Service's trial of live facial recognition technology. Report, The Human Rights Big Data and Technology Project, UK, July. Available at: <https://repository.essex.ac.uk/24946/1/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report-2.pdf>.
- Garland D (2004) Beyond the culture of control. *Critical Review of International Social and Political Philosophy* 7: 160–189.
- Gates KA (2011) *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York: New York University Press.
- Haggerty KD and Ericson RV (2000) The surveillant assemblage. *The British Journal of Sociology* 51(4): 605–622.
- Herzig R (2017) *The role of symbolic capital in digital inequality: Lessons from the student room's reputation system*. PhD Thesis, University of East Anglia, UK.
- Interpol (2018a) Speaker 04. Available at: <https://www.youtube.com/watch?v=foXSJctHSqs> (accessed 18 June 2021).
- Interpol (2018b) Speaker Identification Integrated Project. Available at: <https://www.interpol.int/en/Who-we-are/Legal-framework/Information-communications-and-technology-ICT-law-projects/Speaker-Identification-Integrated-Project-SIIP> (accessed 8 December, 2021).
- Kak A (2020) Introduction. In: Kak A. (ed), *Regulating Biometrics: Global Approaches and Open Questions*, New York: AI Now, pp.62–69.
- Khelif K, Mombrun Y, Backfried G, et al. (2017) Towards a breakthrough speaker identification approach for law enforcement agencies: SIIP. In: 2017 European intelligence and security informatics conference (EISIC), Athens, Greece, 11–13 September, pp.32–39: IEEE.
- Khelif K, Mombrun Y, Hazzani G, et al. (2018) SIIP: An innovative speaker identification approach for law enforcement agencies. In: STO meeting proceedings paper, NATO-OTAN, pp.1–14.
- Kind C (2019) Biometrics and facial recognition technology – where next? In: Ada Lovelace Institute. Available at: <https://www.adalovelaceinstitute.org/blog/biometrics-and-facial-recognition-technology-where-next/> (accessed 4 July 2021).
- Kindt E (2020) A first attempt at regulating biometric data in the European Union. In: Kak A (ed) *Regulating Biometrics: Global Approaches and Open Questions*. New York: AI Now, pp.62–69.
- Kofman A (2018) Interpol rolls out international voice identification database using samples from 192 law enforcement agencies. *The Intercept*, 25 June.
- Leese M (2020) Fixing state vision: Interoperability, biometrics, and identity management in the EU. *Geopolitics*: 1–21. DOI: 10.1080/14650045.2020.1830764.
- Li SZ and Jain AK (2015) *Encyclopedia of Biometrics*. Boston, MA: Springer.
- Li Z, Shi C, Xie Y, et al. (2020) Practical adversarial attacks against speaker recognition systems. In: HotMobile 2020 - proceedings of the 21st international workshop on mobile computing systems and applications, New York, USA, March 2020, pp.9–14: Association for Computing Machinery.
- Liberty (2020) Liberty wins ground-breaking victory against facial recognition tech. In: Liberty. Available at: <https://www.libertyhumanrights.org.uk/issue/liberty-wins-ground-breaking-victory-against-facial-recognition-tech/> (accessed 26 June 2021).
- Lyon D (2008) Biometrics, identification and surveillance. *Bioethics* 22(9): 499–508.
- Madikeri S, Motlicek P and Dey S (2019) A Bayesian approach to inter-task fusion for speaker recognition. In: ICASSP 2019 - 2019 IEEE international conference on acoustics, speech and

- signal processing (ICASSP), Brighton, UK, 12–17 May 2019, pp.5786–5790.
- Marciano A (2019) Reframing biometric surveillance: From a means of inspection to a form of control. *Ethics and Information Technology* 21(2): 127–136.
- McBride C (2013) *Recognition*. Cambridge, Malden: Polity.
- Morrison GS, Sahito FH, Jardine G, et al. (2016) INTERPOL Survey of the use of speaker identification by law enforcement agencies. *Forensic Science International* 263: 92–100.
- Park U and Jain AK (2010) Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security* 5(3): 406–415.
- Poddar A, Sahidullah M and Saha G (2019) Quality measures for speaker verification with short utterances. *Digital Signal Processing* 88: 66–79.
- Pollack I, Pickett JM and Sumbly WH (1954) On the identification of speakers by voice. *Journal of the Acoustical Society of America* 26(3): 403–406.
- Rashid RA, Mahalin NH, Sarijari MA, et al. (2008) Security system using biometric technology: Design and implementation of voice recognition system (VRS). In: 2008 international conference on computer and communication engineering, Kuala Lumpur, Malaysia, 13–15 May 2008, pp.898–902.
- Sánchez-Monedero J and Dencik L (2020) The politics of deceptive borders: ‘biomarkers of deceit’ and the case of iBorderCtrl. *Information, Communication & Society*: 1–18. DOI: 10.1080/1369118X.2020.1792530.
- Scott JC (1998) *Seeing Like A State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven: Yale University Press.
- Taylor C (1994) *Multiculturalism: Examining the Politics of Recognition*. Princeton, NJ: Princeton University Press.
- Taylor L (2017) What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 2017: 1–14.
- Turow J (2021) *The Voice Catchers: How Marketers Listen In to Exploit Your Feelings, Your Privacy, and Your Wallet*. New Haven: Yale University Press.
- Valentino-DeVries J (2020) How the police use facial recognition, and where it falls short. *New York Times*, 12 January.
- Van der Ploeg I (1999) The illegal body: Eurodac’ and the politics of biometric identification. *Ethics and Information Technology* 1(4): 295–302.
- Van Zoonen L (2013) From identity to identification: Fixating the fragmented self. *Media, Culture & Society* 35(1): 44–51.
- Williams P (2015) Criminalising the other: Challenging the race-gang nexus. *Race & Class* 56(3): 18–35.
- Williams P and Clarke B (2016) *Dangerous associations: Joint enterprise, gangs and racism*. Report, Centre for Crime and Justice Studies, London, UK, January.
- Young IM (2011) *Justice and the Politics of Difference*. Princeton and Oxford: Princeton University Press.