Who owns your voice? Linguistic and legal perspectives on the relationship between vocal distinctiveness and the rights of the individual speaker

Dominic Watt, Peter S. Harrison and Lily Cabot-King

University of York

Abstract

Only in very recent times has the concept of 'ownership' of a human voice begun to demand proper consideration in terms of its legal implications. The current lack of clarity with respect to the rights afforded to individuals and organisations in this area is something that must be addressed as a matter of some urgency, given that voice samples are now collected on an unprecedented scale, with or without the knowledge or consent of the person(s) who produced the captured speech. In this paper we explore the issue of voice ownership from a variety of perspectives, starting with an attempt to define what the voice actually is, and then considering how representations of a talker's voice at greater or lesser levels of concreteness (or 'tangibility') can be misappropriated and misused in unethical or unlawful ways.

1. Introduction

Who owns your voice? In all probability most people have never been asked this question, and they may well never have even considered the idea of voice ownership before. The very premises implied by the question – that a voice is something that can be 'owned', and that someone other than speaker X could claim ownership of X's voice – may seem puzzling, and trying to answer the question may appear too whimsical a thought experiment to be worth dedicating any effort to. Such attitudes are hardly surprising. Until relatively recently the issue of voice ownership was one that would scarcely ever have arisen outside of the realm of mythology, fables or folk tales. The Greek myth of Echo and Narcissus, and Hans Christian Andersen's fairy story 'The Little Mermaid', are examples of narratives in which a jealous antagonist maliciously seizes possession of a character's voice as retribution for some perceived offence. No doubt similar stories have existed in many cultures throughout human history, sharing as a common thread the idea that one's voice is an integral element of one's personal identity and concept of selfhood, and that forfeiting its ownership to someone else would inflict deep psychological suffering on the voice's original owner.

It is not just in traditional storytelling that we find evidence of the voice being viewed as an inalienable part of the talker's personal identity, at least to the extent that he or she retains the physical capacity to speak. It is true that under circumstances in which speakers are prohibited from using their native language (e.g. Catalan, Galician and Basque in General Franco's Spain, or Welsh in schools in Wales during the nineteenth and early twentieth centuries) they may feel that they have been 'robbed of their voice', their language being a central element of their sense of self. But it must be rare for anyone to be denied the right to speak altogether, even if that person has been enslaved or forced to take a vow of silence for religious reasons. Only in a figurative sense could anyone be said to have taken 'ownership' of the victim's voice in these scenarios. In a

similar way, individuals who lose the ability to speak because of injury or disease may say they feel that their voice has been stolen from them, but unless the loss of the capability has come about through medical negligence (e.g. a botched surgical operation) or a physical attack (e.g. a deliberately inflicted head or neck injury), it is not possible to apportion blame for the loss of vocal function to any particular agent.

Given recent technological advances, however, the question of voice ownership is one that needs to be confronted in a serious way. Our voices are routinely recorded with or without our consent, and samples of our speech may be being used for commercial or state-sanctioned purposes in ways we are not aware of (Cowling 2015; Bohm et al. 2017). Where voice samples can be obtained illicitly with apparent ease, for example via phishing phone calls, we must also be increasingly vigilant against the misuse of samples of our speech for fraudulent or other criminal purposes (e.g. Chen et al. 2017; Feng, Fawaz & Shin 2017; Shirvanian, Saxena & Mukhopadhyay 2018). We should also consider the issue of voice ownership in terms of its market value, whether in monetary terms or more abstractly in terms of how the misappropriation of a speaker's voice might impact on his or her image or public reputation. This is no longer the concern just of performing artists such as singers, whose living may very heavily depend upon the characteristics of their voices, and to whom the misappropriation of voice samples may bring material losses (see e.g. McLeod & DiCola 2011; Sewell 2014; Ihalainen 2018).

In this paper we take a broad view of the voice as the property of any individual which may be liable to theft or misuse, whether or not that person relies upon his or her voice as a means of earning a livelihood. Our aim is to scrutinise the implications of the voice ownership question first through a linguistic lens, and then from a legal perspective, from which we look at protections offered to individuals in three different jurisdictions: principally England and Wales and particular applications of the law of the United States and of France.

2. What is the voice?

It is interesting to note that linguists find it almost as difficult to give an answer to this deceptively simple question as non-linguists do. Kreiman & Sidtis (2011), who review the very substantial quantity of prior literature on voice, observe that

> [a]lthough a clear definition of voice is a prerequisite to its study, the broad range of functions subserved by voice has made it difficult to provide a single, all-purpose definition that is valid and useful across disciplines, scholarly traditions, and research applications. As voice scientist Johann Sundberg has noted (1987), everyone knows what voice is until they try to pin it down, and several senses of the term are in common use. (2011:5)

For our purposes, we may wish as a start to sketch out a definition of the voice along the lines of 'the set of sounds produced by a normally functioning human vocal tract', where the vocal tract refers to the passage inside the head and neck composed of airfilled cavities lying between the upper trachea and the open air, along with the mobile anatomical structures that can voluntarily be reconfigured so as to alter the shape and size of those cavities. But our working definition encompasses too broad a spectrum of

acoustic phenomena: it would include reflexive or otherwise involuntary sounds produced by the vocal tract such as coughs, sneezes, breath noises, sniffs, hiccups, chewing and crunching noises made as we consume food, swallowing sounds, the transient clicks and popping sounds caused by saliva and the (non-linguistic) movements of the tongue, lips and soft palate, and so forth. However, neither would we necessarily want to confine our definition just to linguistic phonetics, i.e. those sounds which play a role in speech and language production (Ladefoged 1971; Laver 2017). While the segmental components of speech (the vowels and consonants), the voice quality or 'vocal timbre' (the product of phonation characteristics combined with the speaker's long-term articulatory setting), and the prosody (suprasegmental features that rely on changes in pitch and intensity) are obviously the critical elements of the voice, we would not want to rule out some of the sounds that a person's vocal tract makes which are not strictly linguistic, or linguistic at all. These might be features such as cries of alarm, anger, surprise, delight, relief, pain, or distress (e.g. Scherer 1995; Roberts 2011), sighs, laughter sounds, hesitation markers such as filled pauses ('ums' and 'ahs'), sobbing and weeping sounds, hummed tunes, and so on, all of which are manifestations of the voice but which lack any semantic content as such. The sounds may nevertheless on their own be sufficient to reveal to the listener the identity of the person producing them, if the latter is familiar to the former, so to that extent we can consider these vocalisations to feature the acoustic/auditory hallmarks of the producer's voice. Moreover, speakers can produce language-like utterances which are linguistically meaningless, as in the scat talk used by jazz singers such as Cab Calloway, Ella Fitzgerald or Louis Armstrong, nonsensical nursery and playground rhymes, 'mock' discourse for entertainment purposes (e.g. the highly popular 'What English Sounds Like to Non-English Speakers' YouTube videos), or the glossolalia practised by certain religious sects (e.g. Chouiter & Annoni 2018).

We might also want to include sounds that are produced when the vocal tract is not functioning normally. For instance, although the speaker's voice may temporarily sound different for health reasons (denasalised, for example, because s/he has a blocked nose as a result of a heavy cold, or rough/harsh in phonation due to laryngitis), we would not want to say that the speech and non-speech sounds that s/he is producing while in this condition are not part of his or her voice, even if they are untypical of it. In more extreme cases, perhaps because the consequences of disease or injury to the speech organs, we may say that the speaker's voice has changed permanently. In the case of Foreign Accent Syndrome, for instance, damage to those parts of the brain responsible for the phonetic encoding of utterances may result in the subject very abruptly starting to use a voice which is unrecognisable not just to the subject's friends and family but even to him- or herself (Keulen et al. 2017). Other neurological conditions such as Parkinson's Disease may affect the sound of the speaker's voice (e.g. greater levels of vocal tremor, disruption to the organisation of breath groups, etc.; Smith & Caplan 2018), while psychological disorders such as depression are known to be correlated with changes in the speaker's vocal behaviour (e.g. Hashim et al. 2017).

The normal ageing process is responsible for quite radical changes to the speaker's voice quality across the lifespan (e.g. Torre & Barlow 2009; Rhodes 2017). The development of the vocal tract in childhood is obviously responsible for many of the most marked alterations in vocal timbre, and for boys the disproportionate enlargement

of the larynx during puberty means a substantial drop in the pitch of the voice even if the vocal tract itself is otherwise not very dissimilar from that of a girl of the same height and build. In middle age the voices of both men and women will tend to deepen in pitch (Kasuya & Yoshida 2017), while in old age the pitch may start to rise again as the tissues of the laryngeal cartilages and vocal folds become increasingly rigid through ossification (Paulsen et al. 2000). This is before we consider any of the other changes that accrue over the individual's lifetime as a result of behaviour that can damage the delicate edges of the vocal folds – smoking, alcohol consumption, or habitual shouting, for example (e.g. Eckert & Laver 1994) – or sociolinguistically-motivated changes that speakers adopt in response to nonlinguistic pressures such as a desire to conform to societal norms (e.g. Harrington 2006). It is remarkable, given this constellation of age-related changes, that the features of a speaker's voice which make that voice distinct from the speech of other individuals may remain sufficiently stable through the speaker's life as to make him or her recognisable to listeners across quite significant timespans (Sidtis & Kreiman 2012; Wenndt 2016).

We cannot simply cleave to a speaker-centred view, either. Kreiman & Sidtis begin (as we have here) by focussing on speech production. They expand from the very specific meaning of voice in the sense of 'phonation' – sound produced by the vocal folds of the larynx – to a broader definition more or less synonymous with 'speech':

> ... [V]oice in this sense includes the acoustic results of the coordinated action of the respiratory system, tongue, jaw, lips, and soft palate, both with respect to their average values and to the amount and pattern of variability in values over time. (Kreiman & Sidtis 2011:6).

However, they note that when attempting to provide a definition of the voice we should not, indeed cannot, sideline the role of the perceiver. Following Bertau (2008), they contend that speech cannot be divorced from its communicative setting; an observer is needed for speech to be meaningful event, and speech is indexical of the body which produced it. The voice, according to Sidtis & Kreiman (2012:146), is 'an embodiment of self in a social context, contributing to expression, perception and mutual exchange of self, consciousness, inner life, and personhood'. It takes, they argue, 'the whole brain and, by extension, the whole person to participate in producing and perceiving a voice' (2012:147). A voice, in this view, is not just something that a speaker creates in the act of speaking, it is a phenomenon that is brought fully into being by virtue of a listener being present to experience it.

Clearly, then, we are faced with some major conceptual problems when seeking to pin down what we mean when we talk of 'the voice'. By Kreiman and Sidtis's argument, we must not focus solely upon the sounds that emanate from a speaker's vocal tract. We need also in some way to incorporate the subjective percepts of the hearer. After all, the listener is the principal source of auditory judgements about factors such as the similarity of two voices, the distinctiveness of one voice relative to others, and its degree of familiarity, three dimensions which are critically important to the thesis we wish to develop in the current paper. Speaker X's voice is defined, in one sense, by the set of exemplars of what listeners familiar with speaker X would accept as a faithful rendition of his or her voice (what Belin, Fecteau & Bédard (2004) call the 'auditory

face', and what Lavan et al. (2019) label 'telling people together' as opposed to 'telling people apart'). We can imagine scenarios in which the testimony of close friends and relatives concerning the authenticity of an utterance apparently produced by a particular speaker might be of paramount importance, for instance in the context of a police investigation into the disappearance of an individual whose voice had been captured on a CCTV video soundtrack or in a recording of a telephone call.

Even familiar listeners can be fooled by skilled impersonators, however, and there is no shortage of would-be spoofers who are prepared to try to pass themselves off as other individuals for personal gain. For example, in 2010 the Polish footballer Artur Boruc, who at the time was goalkeeper for the Glasgow Celtic football team, became the target of a hoax which involved offering the now defunct British tabloid newspaper News of the World a set of recordings of ostensibly authentic intercepted telephone calls made by Boruc from a Glasgow hotel. It was claimed by the seller that the recordings contained sexually explicit conversations between Boruc and a girlfriend which, given their titillation value, were worth a considerable amount of money. The paper immediately paid the seller a five figure sum for the recordings, but quickly started to suspect that they had been sold a fake. After brief examination of the material by forensic speech analysts, it rapidly became clear that the News of the World had indeed been defrauded. The similarities between Boruc's speech and that of the voice in the recordings were tenuous at best, with the impersonator's efforts to adopt a Polish accent like that of Boruc being confined to making his speech slower and more hesitant, and inserting the occasional Polish word. Boruc could in any event prove that he was on holiday abroad at the time that the calls were made (Greenslade 2011).

The case record contains less amateurishly contrived and considerably graver crimes of impersonation than that described above, however. In 2005 a man from East Yorkshire called Richard Garner murdered his father Frank and then pretended to be his father in telephone calls to Frank's bank so as to transfer funds into his own (Richard's) account. Garner's impersonation temporarily succeeded, in that over the course of several years Garner managed to obtain tens of thousands of pounds in this manner before being caught (Yorkshire Post 2009). Impersonation intended simply as a practical joke can escalate in unexpected ways, in one recent case with tragic results. Australian radio presenters Mel Greig and Michael Christian decided while on air to place a prank call to King Edward VII's Hospital in London, at which the Duchess of Cambridge was being treated for acute morning sickness. The call was taken by an Indian nurse called Jacintha Saldanha, who was duped by Greig's and Christian's live impersonations of Queen Elizabeth II and Prince Charles, thinking that their inquiries after Middleton's health were genuine. Saldanha accordingly transferred the call to the nurse who was taking care of Middleton. Three days after the prank call was broadcast, Saldanha committed suicide (Davies 2014).

In light of the viability of fraud by skilled impersonators, it is only to be expected that occasionally crime suspects who are seeking to exonerate themselves may claim that a voice heard by a witness or in a recording is not their own, but that of an impersonator. A relevant example comes from the case of Tommy Sheridan, former member of the Scottish Parliament, leader of the Scottish Socialist Party and subsequently founder of the Solidarity Party, who was convicted for perjury in 2010. Sheridan had been

suspected of lying under oath during a defamation trial that took place in 2006, and in an effort to prove that he had done so a covert video recording of Sheridan was arranged by one of his associates, George McNeilage. Sheridan cannot be clearly seen in the recording as the camera was all but completely concealed, but the soundtrack of the conversation with McNeilage is intelligible and of reasonably high quality. It contains an admission that Sheridan had lied in the 2006 trial in order to protect himself. When confronted with this evidence, Sheridan denied that the speaker was him; he claimed that although the voice of McNeilage's unseen companion was similar to his own, it must have been either the voice of an impersonator, or a composite made up of genuine samples of Sheridan's own speech spliced together. The former claim is not altogether fanciful: the TV and stage comedian Des McLean is known for his accurate impersonations of Tommy Sheridan, and it is of course conceivable that an equally talented impersonator might have been recruited to frame Sheridan using these methods (Zetterholm 2007; McClelland 2011). Whether the second strategy would have been feasible at the time is harder to determine. Splicing segments of natural utterances together generally results in very unnatural-sounding speech (cf. Chris Morris's 'Bushwhacked' comedy videos, which are mosaics of excerpts of speeches by former US President George W. Bush). For a variety of technical and linguistic reasons it would certainly be challenging to create a soundtrack built from samples of Sheridan's speech in which the junctures between neighbouring samples were inaudible. A good deal of electronic editing would be needed to smooth over the joins. Today, however, we have 'voice cloning' software that might make this task considerably easier. We will return to consider the potential misuse of these tools in section 4.

3. Defining the voice: practical obstacles

For our current purposes it would be highly desirable if we could more clearly delineate what the voice is in a practical sense. After consideration of the issues outlined in the preceding sections, we might offer the following:

> The voice is the output in the sound domain of the coordinated movements of an individual human talker's speech organs, which is used principally but not exclusively by the talker for the acoustic encoding of linguistically meaningful utterances, and which after a period of exposure listeners may come to associate with that talker.

The above definition allows for the fact that not all of the sounds that a speaker uses his/her voice to produce are of linguistic import, and it takes account of both productive and perceptual aspects. But how can we circumscribe the voice for our current purposes? Is the voice of speaker X something that can be abstracted from the sum total of all the utterances that a speaker has ever said, or could say? What parameters would we choose to quantify so as to achieve this? A parsimonious representation is presumably preferable as long as it is sufficient to distinguish speaker X from every other speaker, but given the impossibility of sampling the speech of every speaker in existence, how would we know when to stop? As yet we do not know how much detail is needed to define a human voice uniquely; indeed, we do not actually know for certain whether human voices are unique to individuals, in spite of how commonly this is assumed (cf. the analogous situation for fingerprints; e.g. McGettigan & Lavan 2017).

We do know that there are talkers – often siblings – who are justifiably described as 'vocal twins' in that they are effectively indistinguishable from one another by voice. These soundalikes need not be genetic twins, and indeed they do not necessarily have to closely resemble one another anatomically either. We note that even monozygotic twins are never exactly identical anatomically, and two anatomically near-identical speakers might in any case put their vocal organs to use in markedly different ways, making them sound different from one another in spite of their physical resemblance (Loakes 2006; San Segundo & Mompean 2017).

In any case, if physical records are lacking we only have listeners' subjective judgements to rely upon. Perhaps we could set about capturing speaker X's voice as the aggregation of acoustic-phonetic properties that can be extracted and averaged from all of the speaker's utterances that have ever been recorded. This would eliminate a large measure of subjectivity, and it would mean that multiple observers could derive a model of the speaker from the same corpus of recorded material. However, given the way that recording technology works we must accept that a recording of someone's voice is to that voice what a digital photograph is to the appearance of the person being represented (see further Section 5.2.1, below). There are many limitations at work here: the quality of the recording equipment, the duration of the speech samples, technical issues to do with sampling rate, bit depth, microphone distance, etc., the channel (e.g. telephonic speech versus studio-quality samples versus open-field recordings), as well as the nature of the captured utterances (spontaneous speech versus read speech, and so on). Ultimately, our model of speaker X's voice would be constrained by the quantity and quality of the material available, and once again by the parameters we choose to use to build the model. Even now, after many decades of research into what makes human voices distinct from one another, it is not fully evident what those parameters would be (Nolan, McDougall & Hudson 2011; Lavan et al. 2019). The solution here may be to remove the human observer as far as possible, and to hand the task of determining what makes speakers sound similar to/different from one another over to machines which have been specifically designed for the purpose.

4. Computer-based models

At present, probably the closest we can come to the objective representations we seek are the models generated by Automatic Speaker Recognition (ASR) systems of the sort now routinely used by police forces and intelligence agencies around the world (Gold & French 2011; Morrison et al. 2016; Watt & Brown 2019). These systems compare digital voice samples in a way that does not resemble human speech perception very closely. Instead, they abstract speaker models based on parameters known as 'Mel-frequency cepstral coefficients' (MFCCs; Mermelstein 1976). The technique computes a 'cepstrum' – informally, a 'spectrum of the acoustic spectrum' – of short-duration (c. 20 millisecond) windows of the digital signal, and concatenates these frames to produce a speaker model characterised by a small number of coefficients coarsely describing the most prominent spectral components of the signal. The MFCC approach has proven extremely successful for speech recognition applications (what is being said) and is also demonstrating its worth in the area of speaker recognition (who is saying it). In principle, the speaker reference models or 'voiceprints' generated using ASR systems such as the market-leading Nuance Forensics package (Nuance Communications 2018)

are effectively unique to individual talkers.[1] The idea is that although speaker models that are generated from

speaker X's speech samples are going to be somewhat different from sample to sample even where it is known beyond doubt that the samples were produced by the same speaker, they will be more similar to one another than they are to models computed for other speakers. To borrow Lavan et al. (2019)'s terminology, these systems are designed to tell people together, as well as to tell them apart.

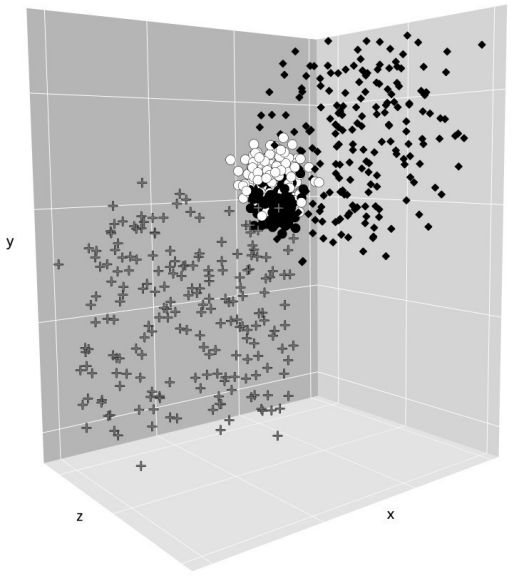

Figure 1. Imaginary multidimensional speaker space in which samples of an individual talker's speech occupy portions of space overlapping to a greater or lesser extent depending on their degree of vocal similarity with samples for other speakers (four speakers in total are represented here). The location of each cloud of points is determined by the acoustic-phonetic properties of each speaker's voice, while the volume of each speaker's space varies according to how variable his/her speech is with respect to these acoustic-phonetic parameters.

We might visualise speakers' voices as occupying a particular region of a multidimensional space, in which individuals are represented by clouds of points corresponding to features of their voices (Figure 1). Speakers' clouds will of course overlap with one another: all speakers will – by definition – bear some phonetic resemblance to other people. But even supposing that speakers may be classed as vocal twins (triplets, etc.) they will nevertheless be found to be distinct from their closest vocal match(es) if we measure properties their speech at a sufficiently detailed

technician Lawrence Kersta (1962, 1966), who seized the opportunity to train operators

[1] The term 'voiceprint' is considered problematic by forensic speech scientists in light of the history of its use, initially by the researchers who developed the analogue spectrograph at Bell Labs (Potter, Kopp & Green 1947), and later by Bell Labs

in the use of the commercialised form of the spectrograph (the 'Kay Sonagraph'; see also Tosi et al. 1972). Kersta's advocacy of visual comparison of printed spectrograms of short speech samples for forensic purposes was heavily criticised at the time, and his techniques have long been considered unsound by the scientific community (e.g. Bolt et al. 1970; Hollien 1990). Nonetheless, as far as the general public is concerned – as well as intelligence agencies and members of the judiciary, in some countries – the concept of the voiceprint seems to be a perfectly acceptable one. We follow Nuance's use of the (admittedly transparent) term for the sake of convenience, but with reservations that are rooted in the discrediting of Kersta's work in the 1960s and '70s. level, at least so long as we accept the hypothesis that the speech patterns of no two speakers are absolutely phonetically identical.

The notion of speaker uniqueness is considered sufficiently established that speech technology companies and their client firms present it to customers as an uncontroversial given. Systems like those currently implemented for telephone banking applications are claimed to be all but infallible. Take, for example, the assurances given by HSBC for the technology underlying their 'Voice ID' system:

> Like your fingerprint, your voice is unique, which means you can create your own voiceprint. Our Voice ID technology analyses your voice in seconds. It checks over 100 behavioral and physical vocal traits, including the size and shape of your mouth, how fast you talk and how you emphasise words.
>
> Once you've created your voiceprint, you'll be able to use your voice to access telephone banking. It also means we can use it to help protect you further against fraud.
>
> Fraudsters and hackers may be able to steal or guess your security number, but they can't replicate your voice. Voice ID is sensitive enough to detect if someone is impersonating you or playing a recording. It can even recognise your voice if you have a cold or a sore throat. (HSBC UK 2019)

For all the confidence with which these claims are expressed, it has been shown that spoofing the HSBC system – state-of-the-art secure biometric technology, in which the corporation has invested many millions of dollars – is far from impossible, and it does not require any high technology. In a widely-publicised story from May 2017 it was reported that Joe Simmons, the (non-identical) twin brother of BBC journalist Dan Simmons, had vocally impersonated his brother Dan while trying to access his (Dan's) HSBC bank account using the Voice ID interface, and had succeeded in gaining entry to the account on only his eighth attempt (Simmons 2017). On learning of this, HSBC immediately pledged to tighten the security of the system by making the criteria for matching genuine customers' speech to the stored voiceprint more stringent. But raising the bar in this way comes at the cost of user-friendliness: customers are unlikely to react well if it becomes more difficult to legitimately access telephone banking facilities using spoken passphrases, even if they accept that their accounts are more secure as a result (ironically, objectors may be more comfortable using the familiar handwritten signature, which can hardly be said to be unforgeable; see further Mukhopadhyay, Shirvanian & Saxena 2015; Sanjith 2017).

The introduction of any purportedly secure gatekeeping technology also brings with it the risk of new kinds of spoofing attacks using artificial techniques to mimic biometric data. Within the last two to three years, it has become possible to electronically emulate individual speakers' voices with a high level of realism. Socalled 'voice cloning' software, such as the Adobe VoCo, Lyrebird, CandyVoice, CereVoice Me, or Baidu's Deep Voice packages (Ping et al. 2018), has a variety of benign applications, for example to assist patients who have lost the ability to speak because of degenerative illness (e.g. Veaux, Yamagishi & King 2015). But it poses a threat if it can even occasionally successfully gain access to secure data – financial data, personal information, medical records, etc. – or to controlled spaces and buildings (military installations, banks, hospitals, etc.). Voice cloning systems use samples of a talker's speech to construct a bespoke speaker reference model, which serves as the template for wholly novel synthetic utterances generated via a text-tospeech interface, i.e. typed messages which the machine then speaks aloud. Although the products listed above have only recently become available, the results are certainly good enough to lead listeners to believe that they are hearing a natural voice rather than a synthesised one (e.g. Wu et al. 2015; Wester, Wu & Yamagishi 2015), and they can at times replicate a voice to a sufficiently high degree of accuracy that they can fool human listeners into thinking they are hearing the authentic speech of a familiar talker (Singh, Jiménez & Øland 2017). It is possible that because of the ubiquity of mobile telephony and VoIP (Voice over Internet Protocol) media such as Skype, WhatsApp or Facetime, listeners are more tolerant than they once were of distortion and glitches in the speech signal because they tend to attribute these to temporary technical issues affecting the channel, rather than to properties of the voice itself. Synthetic speech that might once have sounded artificial may now more often be assumed to be genuine. Further research on this question would be valuable.

In the case of HSBC and other organisations making similar use of speaker verification technology, the speech samples donated by customers take the form of a small number of repetitions of the short passphrase 'My voice is my password', transmitted via the telephone.[2] While at present this kind of passphrase-based access is only considered secure enough to allow the customer to perform tasks such as checking a balance or transferring funds to another account registered in his/her name with the same bank, the list of applications of the technology is likely to expand as consumers become more accustomed to, and more confident with, the routine use of speaker verification (cf. the use of contactless credit and debit cards; Akinyokun & Teague 2017). A concurrent rise in the incidence of identity theft, and the identity fraud that it enables, is therefore to be expected. Recent figures published by the Crime Survey for England and Wales (2018) show that the rate of 'remote banking fraud' (criminal misuse of telephone and internet banking) has followed an upward trend over the past five years or so, and although it

[2] In May 2019, HMRC (Her Majesty's Revenue and Customs, the UK central tax office) was forced to delete 5 million voice files after the Information Commissioner's Office determined that they had been collected from taxpayers in a way that broke privacy rules. HMRC will continue to use the same voice biometrics security system, however, having pledged to obtain explicit consent from users before recording their passphrases (see Peachey 2019).

seems to have declined slightly over the course of 2019 it is also recognised that reporting rates are alarmingly low. The 2017 Survey estimated that just 17% of cybercrime attacks committed via the telephone were reported to the authorities (Crime Survey for England and Wales 2017). These include phishing attacks, in which a caller – whether human or automated – will typically seek to extract personal or confidential information from the call recipient. Moreover, given the potential of voice cloning software to create entirely novel utterances on the basis of the speech of the victim of such an attack, there is further scope for fraudulent misuse by other parties if the target's voice samples are obtainable by directly hacking the speaker verification system itself (Mirra 2018). Alternatively, samples of speaker X's voice might be obtained through a breach of the secure storage of instructions or questions addressed to the now widely-used speech/speaker recognition systems Alexa, Google Home, Siri and

Cortana, which are built into an increasing variety of familiar electronic devices such as smartphones, domestic smart speakers, and in-car satnav systems, as well as commercial and governmental chatbots and more specialised equipment intended for hands-free use in, for example, operating theatres or to control military hardware. The fact that these technologies are designed to learn differences between individual speakers who might make use of them (for instance, multiple members of a single household issuing commands to a smart speaker), without needing to be explicitly trained to, indicates that person-specific voice data is being generated on a gigantic scale. On the face of it, it seems implausible that these records should be more difficult to illicitly obtain than other sorts of personal data. The recent history of major data breaches perpetrated by hackers, or via accidental release into the public domain, may with justification lead consumers to mistrust the large tech companies (e.g. Armerding 2018; Jones 2018). It may therefore be that estimates of the future uptake of voice biometrics-based security systems, such as the projection by Miller (2016) to the effect that by 2020 some 500 million customers worldwide will have enrolled their voiceprints for telephone banking applications, are over-inflated. It is likely, nonetheless, that the numbers of customers will continue to rise from their current levels, which according to Nuance Communications had already exceeded 300 million by the end of 2017.[3]

5. Ownership

We return now to our original question. To whom does a voice, or indeed voice samples, belong? What can I do to stop others from using my voice or voice samples? We will see that there is no one law that governs ownership of the voice. Instead there is a 'matrix' of many rights which do, to varying extents and in varying ways, give some degree of control. We will also see that control will very much depend on what a third party is doing with a voice and the reason they are doing it. In some cases we will see that there are marked limits to what the law can do.

5.1. The problem of 'fuzzy edges' and tangibility

[3] Source: <https://www.nuance.com/en-gb/omni-channel-customerengagement/security/whats-new.html>, accessed 5th May 2019.

The concept of 'property' is at the heart of preventing others from doing something with 'things' which are (or at least you believe should be) yours. The law has a long history of protecting property and determining what constitutes (or should constitute) property, who should own a particular property and what you can and cannot do with another's property. We have got used to the idea of property rights in tangible items - things which you can touch. I can intuitively understand how I might lock away my Vermeer painting, or fence off my Scottish Highland deer forest. There are many questions as to whether I should have the right to keep my Vermeer away from public appreciation, or prevent others from utilising my otherwise rarely-used tracts of wilderness. For both, however, the thing which is the subject of legal protection can be easily defined. Land lawyers, and land registries, will delineate land by the drawing of a coloured line on a map around the land over which ownership is claimed. However, when we come to look at ownership over an intangible thing the certainty provided by a coloured line disappears.

As an intangible 'thing' cannot be physically locked away or fenced off, it can only be protected through the application of positive law. If we are to do that, we need to establish a way in which we can define the legal limits of protection of the intangible item, such that third parties understand what is, and what is not, owned. As I cannot simply draw around my intangible concept with a red coloured pencil, I need to somehow describe it in sufficient detail in what way my intangible 'thing' is different from all other such 'things'. That is of course easy to say, but as soon as one begins the exercise one sees that are many subjective judgments to be made.

Intellectual property laws seek to protect a wide range of intangible concepts of many different types (for example by way of patents, trade marks, design rights, copyright, and rights in performance). However, in each case the law must distinguish the intangible from similar existing things (what we will call here backward-facing determinations) and similar things that will be created in the future (forward-facing determinations). In making these determinations the law has to take into consideration the reason for the existence of the right in the first place. Accordingly, in each case the definition of what is to be protected covered is always purposive.

What then of the voice? How might the law go about defining the voice in terms of a thing that can be controlled or 'owned'? There are two key elements. The first is the question of vocal distinctiveness – as discussed above, there are many ways in which we might describe a voice as being different from another and how we might explain why my individual voice is 'mine'. For our legal analysis these reasons for distinctiveness can be defined as objective differences (for example the raw sounds emanating from the vocal tract) and subjective differences, where the determination of difference depends on the perception and opinion of the listener. As we will see, how the law treats these different ways of distinguishing voices is significantly driven by the purpose for protection.

Alongside this objective/subjective split we additionally have a problem of tangibility. Although intellectual property rights broadly deal with intangible 'things', intellectual property lawyers appreciate that there is often a range of relative intangibility within

the things which intellectual property rights seek to protect. For example, a patent will protect an inventive concept which may brought into effect in many ways. In contrast, copyright law broadly protects a particular expression of a concept, rather than protecting a concept per se. Such an expression is arguably somewhat less intangible (having been 'captured' in a medium) than is a more amorphous concept.

Though all elements of a voice are strictly speaking intangible, some are arguably more intangible than others and sit at different places on a spectrum of tangibility. This tangibility spectrum provides a convenient scheme by which to present the various elements of vocal distinctiveness and how they are (or could be) protected. For the present legal analysis we have divided the voice into three levels of relative 'intangibility':

i) The 'captured' voice, where a voice has been captured by an analogue or digital recording;
ii) Voiceprints and computer-based models; and
iii) The 'uncaptured' voice, where the sought protection is not over any particular recording but of a voice in general.

The focus of our analysis will be upon the law of the United Kingdom (or where it is the source of UK law, that of the European Union) but we will also be using examples from other jurisdictions, notably the United States and France, where the law of those jurisdictions may bring particular insights. However, our aim is to look at the principles of ownership over vocal distinctiveness which are of broader application, regardless of jurisdiction.

5.2. The 'captured' voice

5.2.1 Copyright

At arguably the more tangible end of our tangibility 'spectrum' is a voice which has been 'captured' in a fixed form through a recording. This is a technology which has been around since the late nineteenth century and the law has (albeit slowly) evolved to be comfortable with protecting such recordings through the law of copyright. By 1961 the International Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations (the 'Rome Convention') provided international recognition for copyright in sound recordings, including those of the spoken and singing voice. Through the 1971 Convention for the Protection of Producers of Phonograms Against Unauthorized Duplication of Their Phonograms, (the 'Geneva Phonograms Convention')[4] rights holders were given an international right to block imports of counterfeit music recordings and their sale. Accordingly, the majority of jurisdictions worldwide provide protection for sound recordings and have progressively amended their legislation to cover all types of recording and replay, regardless of technology (see, for example, the UK Copyright Designs and Patents Act 1988 ('CDPA') Section 5A). So where a voice has been 'fixed' in a recording, whether in analogue form on a wax cylinder or vinyl disc, on magnetic tape, or as a digital file held

[4] See <https://bit.ly/2DCF78R>.

in whatever way, that recording is protected from being copied. However, it should be noted that the recording will be owned by the person making the recording, not the person who possesses the recorded voice.[5]

However, what of a voice alone? Does copyright protect a voice which has not been 'fixed' in a recording? What does copyright law tell us about the legal protection of vocal distinctiveness?

For these purposes, a recording of a voice is analogous to the taking of a photograph of a face. The photograph captures parts of a face at any particular time, but it is of course not the face as such. Similarly, a recording is only a capturing of a voice at a particular time rather than the voice in the round. Arguably it is the act of 'fixing' the

---

voice in a recording which is what creates copyright in a work, much as copyright in a photograph is created by the taking of photograph. Since copyright is taken to subsist on the creation of a work (rather than as a result of some registration process), the concept of 'fixation' within a tangible medium is often assumed to be key to the creation of copyright.

In Midler v Ford Motor Co. (849 F.2d 460 – 9th Cir. 1988) the United States Court of Appeals was asked to consider a case in which the Ford Motor Company's advertising agency created a series of advertisements which used various classic 1970s songs sung by the original artists. When Bette Midler refused to give permission for her recording of Do You Want to Dance?, the agency sought out Ula Hedwig, a member of Midler's group of backing singers, The Harlettes. Hedwig was asked to sing in a style as close as possible to that of Midler. The court heard evidence that following the launch of the advertisement many listeners believed that it was Midler they were hearing. As the agency had a licence to use the copyright in the song itself from the copyright holder, Midler could only bring a claim for the protection of her voice. However, as regards copyright protection the court was clear:

> Copyright protects 'original works of authorship fixed in any tangible medium of expression.' 17 U.S.C. Sec. 102(a). A voice is not copyrightable. The sounds are not 'fixed.' What is put forward as protectable here is more personal than any work of authorship. (Paragraph 9).

However, as we will see below, the court believed that Midler was entitled to an alternative form of protection under a 'right in publicity'.

---

[5] Many jurisdictions provide a separate 'right in performance' to control the recording of a live performance. For example, Part II of the UK CDPA provides that performers have a right to control a sound or film recording of a dramatic performance and control third parties dealing with illicit recordings. See for example Experience Hendrix LLC v Purple Haze Records Ltd and Ors [2007] UKSC. Performing artists also have a 'moral right' to prevent 'derogatory treatment' of their performance (CDPA Section 205F) and the right to be identified as the performer of a live performance or on a recording of their live performance (CDPA Section 205C).

Not all jurisdictions are as clear on the question of fixation as is the situation in the United States. Indeed, the international agreement governing the mutual recognition of copyright, the Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886 as amended (the 'Berne Convention'), states at Article 2(2):

> It shall, however, be a matter for legislation in the countries of the [Berne] Union to prescribe that works in general or any specified categories of works shall not be protected unless they have been fixed in some material form.

UK law states that copyright does 'not subsist in a literary, dramatic or musical work unless and until it is recorded, in writing or otherwise' (see UK CDPA Section 3(2)). However, although fixation is not expressly required in the UK for copyright to exist in other types of work, it is often implicit in the wording of other sections of the Act that some form of fixation is required, as is certainly the case in respect of sound recordings where it is implicit that there is a recording.

If some type of fixation is the key to the existence of copyright in a voice recording, it arguably tells us nothing in relation to the question of the protection of vocal distinctiveness. It is not the distinctiveness of the voice which creates the right, but the mechanical or electronic process of fixation.

This is hardly surprising. The purpose of copyright is to protect created 'works' from being copied by others so as to reward the authors of the works for their creativity and to incentivise others to put their creative efforts and often financial resources into creating works which will benefit the cultural richness, and sometimes economic wealth, of society. Although for some actors and singers their voice may be a part of their artistic persona and may contribute to the cultural value of their created works, their voice is not a 'work' in itself.

5.2.2. Trade mark protection

There is often a great deal of confusion over the use and appropriateness of trade marks. Lay people often say that someone has a trademark 'style' or 'characteristic' or even that someone has a 'trademark voice'. This often betrays a misunderstanding of what trade marks are intended to (and indeed can) do. We will come to the protection through trade marks of an 'uncaptured' voice alone later in this work (section 5.4), but first we ask whether a distinctive voice captured through a recording can be protected through trade mark law.

A trade mark can be most simply be defined as a recognisable sign, design, or expression which enables consumers to identify a product or service coming from a particular source and to distinguish it from those of other origin. A registered trade mark gives its holder a powerful monopoly right to prevent the use by others of the trade mark for those goods and/or services for which it is registered (and in some cases the use of similar or non-similar goods and services). Given the power of these rights, an application for a trade mark has to meet stringent validity requirements before it is granted. For our current analysis there are two key validity requirements which impact

on the registrability of a particular voice. The first concerns the inherent nature of a voice and the second concerns questions of distinctiveness.

Historically, trade marks were either a word alone, the representation of a word in a stylised font, or a logo. However, with the growth in inventiveness of product promotion, trade mark law has evolved to encompass a number of so-called 'unconventional' trade marks. The majority of unconventional trade marks are visible signs (e.g. colours, shapes, moving images). Classic examples are the distinctive shape of the Coca Cola bottle, the green and yellow colour scheme of John Deere tractors, and the triangular Toblerone chocolate bar. However, in theory at least, unconventional trade marks can extend to extend non-visible signs (e.g. sounds, scents, tastes, textures). Until October 2017, European law (Article 2 of Council Directive 89/104/EEC) required that a valid trade mark had to be capable of being represented graphically and be capable of distinguishing goods or services of one undertaking from those of other undertakings. This graphical representation requirement created problems in relation to describing non-visible signs. The problem can be seen in the decision of the European Court of Justice ('ECJ') in Ralf Sieckmann v Deutsches Patent und Markenamt (Case C-273/00, December 2002).

The applicant Sieckmann had a 'methyl cinnamate' scent which they described in their application for a trade mark as 'balsamically fruity with a slight hint of cinnamon'. The ECJ ruled (in paragraphs 69 to 70) that: (a) a 'chemical formula does not represent the odour of a substance... nor is it sufficiently clear and precise'; (b) a written 'description of an odour, though graphic, ...is not sufficiently clear, precise and objective'; and (c) a physical 'deposit of a sample of the odour ..does not constitute a graphic representation', and is 'not sufficiently stable or durable'.
The Court further held (at paragraph 73) that none of these representations, whether alone or in combination, could satisfy the graphic representation requirement.

The doctrine in Sieckmann was developed and applied to sound marks by the EU Court of Justice in the case of Shield Mark BV v Joost Kist h.o.d.n. MEMEX (Case C283/01, November 2003). Shield Mark were the applicants for a number of sound marks including: 'a musical stave with the first notes of the musical composition Für Elise; others of 'the first nine notes of "Für Elise"'; and others of the sequence of musical notes 'E, D♯, E, D♯, E, B, D, C, A'. Other applied-for marks consisted of the word 'Kukelekuuuuu' (onomatopoeia suggesting, in Dutch, a cockcrow) or of 'a cockcrow''.[6]

The Court considered that sound signs were not expressly precluded from Directive 89/104/EEC, but that 'a sound sign must firstly make it possible to distinguish the goods or services of one undertaking from those of other undertakings'. Secondly, it must be 'capable of graphical representation, in particular by means of figures, lines or characters that are clear, precise, self-contained, easily accessible, intelligible, durable and objective'. It was further held that:

> Those requirements are not satisfied by a graphical representation of the sound sign consisting in the indication that the sign consists of the notes making up a

[6] Curia Press Release 106/03 re Shield Mark BV v Joost Kist h.o.d.n. MEMEX (Case C-283/01, November 2003)

> well-known work or in a simple sequence of musical notes, without more, or again by a graphical representation that is merely onomatopoeia. In that case the graphical representation at the least lacks precision and clarity.
>
> On the other hand, those requirements are satisfied if the sound sign is represented graphically by a musical stave divided into bars and showing a clef, notes and other musical symbols. This notation, as a whole, constitutes a faithful representation of the sequence of sounds which form the melody in respect of which registration is sought.

Following Sieckmann, the Board of Appeal of the EU trade mark registry rejected an application made by the Metro-Goldwyn-Mayer Lion Corporation (MGM) for the sound of a lion's roar, a well-known part of MGM's film introductions since the late 1920s. (Metro-Goldwyn-Mayer Lion Corporation's Appeal relating to Community trade mark application No. 143 891, Case R 781/1999-4). The graphic representation of the roar consisted of a 'sonogram' (i.e., a spectrogram) with text describing 'the sound produced by the roar of a lion'. It was accepted by the Board that a sound sign could be graphically represented using a spectrogram depicting pitch and volume, but on the facts considered that MGM's lion's roar spectrogram was deficient. However, the potential use of spectrograms was later restricted in the case of Edgar Rice Burroughs, Inc. v OHIM (Case R 708/2006/4, September 2007). This concerned a series of applications for Tarzan's distinctive 'yell'. The applications were filed as spectrograms along with a written description describing the Tarzan yell:

> the yell of the fictional character TARZAN, the yell consisting of five distinct phases, namely sustain, followed by ululation, followed by sustain but at a higher frequency, followed by ululation, followed by sustain at the starting frequency.

The Board of Appeal of the EU trade mark registry rejected the applications because they considered that a spectrogram, 'whether on its own or accompanied by a written description', was insufficient to meet the criteria set out in Sieckmann. They held that a spectrogram 'did not permit third parties to reproduce the sound without technical means'. Nobody could 'read' a spectrogram, they contended, and therefore it was not self-contained, clear, intelligible, or easily accessible.

Following these decisions, the practice of the EU trade mark registry was amended (by way of Commission Regulation (EC) No. 1041/2005) such that digital sound files could be submitted along with a with a graphic representation of the sound – viz., a spectrogram – within an electronic filing for a Community Trade Mark. As a result of this, sound marks have been granted for both the MGM lion's roar and the Tarzan yell.

Following an extensive review of EU trade mark law by the EU Commission, the European Parliament approved the EU Trade Mark Regulation (2015/24/24 and the EU Trade Mark Regulation (2015/2436)). The effect of both is to replace the old graphical representation requirement with a requirement merely that a valid trade mark must be capable: '(a) of being represented in the register in a manner which enables the registrar and other competent authorities and the public to determine the clear and precise subject matter of the protection afforded to the proprietor; and (b) of distinguishing goods or services of one undertaking from those of other undertakings.' Accordingly, the EU trade mark registry have amended their practice from requiring (i) musical notation

(with optional sound file) or (ii) a sonogram, accompanied by a sound file, to requiring the submission of an audio file reproducing the sound or by an accurate representation of the sound in musical notation.

It is clear from Sieckmann and following decisions – and the practice of the EU trade mark registry which are followed by national registries of EEA member states – how much importance was placed on a putative trade mark clearly setting out the exact borders of protection provided to the right holder. Although the requirement for graphical representation is no longer present, and the new law would seem to have fewer formal limitations, it is clear that at its core there remains a requirement that all involved, including third parties, can determine the clear and precise subject matter of the protection afforded to the proprietor.

It is possible that an applicant seeking to register a distinctive voice saying a distinctive phrase in a distinctive manner could achieve a successful trade mark application. Submitting an audio file would likely meet the 'clear and precise subject matter' test. The problem that such an application would encounter is that the voice, phrase and manner would have to be capable of distinguishing the goods or services of one undertaking from those of other undertakings. Would the voice element of that application alone be sufficient to provide distinctiveness? That is not clear. We will discuss the impact of vocal distinctiveness on product goodwill where we discuss protection of the 'uncaptured' voice through the law of passing-off (section 5.4.2).

5.3. Voiceprints and computer-based models

In section 4 we discussed Automatic Speaker Recognition (ASR) systems and the creation of 'voiceprints', which are in principle unique to an individual speaker. But who owns such a 'voiceprint'? Are they indeed owned at all?

5.3.1. Copyright protection

If the ASR uses a number of digital voice samples these samples could, in principle, attract copyright protection as voice recordings. However, copyright law across jurisdictions recognises a de minimis principle that if a work is too small, or trifling, it will not be granted protection. Where this de minimis 'line' sits has been hugely problematic for courts in many jurisdictions. In Infopaq International A/S v Danske Dagblades Forening ([19 July 2009] ECR I-6569) the Court of Justice of the EU held that unauthorised use of eleven word long newspaper article extracts amounted to reproduction of a copyright work, provided that the elements which were reproduced were the 'expression of the author's intellectual creation.' Again, what that means is problematic, but it might reflect the concept that if a work is identifiable as a distinct work then it should attract copyright protection. For these authors short voice samples are likely to fall foul of the de minimis rule and the more that a voice sample is fractionated the still less likely it is to attract copyright protection. Even if such a sample did attract such protection, it is important to note that the ownership of the copyright would sit with the organisation that had made the recording of the voice, rather than with the person from whom the samples were taken.

5.3.2. Fraud

What about where voice samples are used for voice cloning? Do I have a right as the person from whom the samples were taken to stop their use? The first point to note is that where a voice is used for fraudulent purposes – for example to gain access to a bank account – then that action will very likely constitute the crime of fraud. In the UK this crime is covered by the Fraud Act 2006. It is a crime under Section 2 of the Act to dishonestly make a false representation, and intend, by making the representation, to make a gain for yourself or another, or to cause loss to another, or to expose another to a risk of loss. A representation is false if it is 'untrue or misleading, and the person making it knows that it is, or might be, untrue or misleading'. 'Representation' here means 'any representation as to fact or law, including a representation as to the state of mind of the person making the representation, or any other person'. Crucially for our purposes: 'a representation may be regarded as made if it (or anything implying it) is submitted in any form to any system or device designed to receive, convey or respond to communications, with or without human intervention'.[7] Clearly, use of voice cloning which constitutes a false representation as to the identity of the person seeking to access a system will (all other elements being present) constitute a fraudulent act. How much vocal similarity is required here? At first blush, it would appear to be similarity that is close enough to achieve the fraudulent end. However, the section is not about the end result but the intention of the fraudster. Attempted use of an imperfect voiceprint where the fraudster did not know of the imperfection, or thought that the imperfection would not be significant enough to prevent successful entry, would still constitute fraud.

5.3.3. Fraudulent misrepresentation

Outside of criminal law the fraudulent use of a voiceprint (or digital spoofing) could give rise to a civil claim under the tort of fraudulent misrepresentation (under the UK

Misrepresentation Act 1967 and related case law). This is where an untrue statement of fact or law made by one party – 'Party A' (or its agent) – to another party, 'Party B', induces Party B to enter into a contract with Party A which thereby causes Party B loss. For fraudulent misrepresentation to apply, the false representation by Party A has to have been made with knowledge as to the untruth of the statement, or without belief in truth of the statement, or recklessly as to truth of the statement. Party B's remedies are rescission of the contract (where the contract is cancelled and the parties put back to where they were before the contact was entered into) and damages arising out of having entered into the contract. It should be noted that this tort is likely to protect the recipient of the fraudulent voiceprint, rather than the person whose voiceprint is being copied. However, neither the crime of fraud, nor the tort of fraudulent misrepresentation, create any sort of 'ownership' in a voiceprint by the person from whom the voice print was created.

5.3.4. Confidentiality/privacy

Some people might consider that a voiceprint created from a sample of their voice might be 'private' to them. Does that sense of privacy give rise to any type of legal ownership

[7] Fraud Act 2006 s2(5)

or control? The answer is not clear. In England and Wales there is no common law right to privacy as such (see Kaye v Robertson ([1991] FSR 62)). However, the law in England and Wales has long protected confidential information from improper disclosure. In the case of Coco v A N Clark (Engineers) Ltd ([1969] RPC 41) the judge stated that three elements would normally be required if a case of breach of confidence is to succeed:

> First, the information itself… must 'have the necessary quality of confidence about it.'[8] Secondly, that information must have been imparted in circumstances importing an obligation of confidence. Thirdly, there must be an unauthorised use of that information to the detriment of the party communicating it.

Claimants have often used this right of action to protect commercially sensitive information, such as client lists and trade secrets, but also to protect information which is of personal significance, such as information relating to their health. If I talk during a phishing phone call to a fraudster who is seeking to record my voice for the purposes of creating a voiceprint so as to later access my accounts, does that constitute disclosure of anything which could be called 'confidential'? There is no confidential 'information' conveyed as such. Neither is my voice strictly confidential.
Under practically all other circumstances I am happy to use it to speak to other people.

In Campbell v Mirror Group Newspapers Ltd ([2004] UKHL 22) the UK House of Lords examined the English law of breach of confidence in the light of the UK's compliance with the European Convention on Human Rights. Article 8 of that Convention requires that 'Everyone has the right to respect for his private and family life, his home and his correspondence.' The court held that, although Article 8 did not create any new cause of action between private persons, if a relevant cause of action existed (such as an action for breach of confidence) the court must seek to apply the

---

principles within Article 8 when examining the existence of a right of action. For Naomi Campbell and her eventually successful case against the Daily Mirror relating to photographs of her leaving a rehabilitation clinic, this meant that her right to expect her private life to be respected needed to be considered. It is hard to know whether a court would consider Article 8 as applying to the situation where I am unknowingly having my voice recorded to create a voiceprint. Even though my voice may not otherwise be of a confidential nature, it is conceivable that (following Campbell) a court might consider that the need for respect of my private life extends a cloak of 'confidentiality' over my voice. We should note that even though this would give rise to a form of control over my voice, that control would be limited to this particular situation. It does not create any right of control over my voice per se, nor is my right of control in any way dependent upon my vocal distinctiveness.

5.3.5. Data Protection

---

[8] The judge, Mr. Justice Megarry, is quoting here from Lord Greene in another case, that of Saltman Engineering Co. Ltd. v Campbell Engineering Co. Ltd. [1948] 65 RPC 203.

The holding and processing of personal information by organisations has progressively become a major concern for the people whose data is being held and legislatures are developing laws controlling the way in which such data can be held and processed. The EU General Data Protection Regulation (EU 2016/679) ('GDPR') imposes a stringent set of rules on those holding personal data. Under GDPR Article 9 even more stringent rules are applied to so-called 'special categories' of personal data which are deemed 'particularly sensitive in relation to fundamental rights and freedoms' (GDPR Recital 51). Such special categories include '... biometric data for the purpose of uniquely identifying a natural person'. Biometric data is defined in GDPR Article 4 to mean:

> ... personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic [fingerprint] data.

Processing of Article 9 data is prohibited unless one of the following applies:

- The data subject has given explicit consent (Article 9(2)(a));
- Processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller or of the data subject in the fields of employment and social security and social protection law;
- Processing is necessary to protect the vital interests of the data subject;
- Processing is necessary for the establishment and exercise of defence of legal claims; or
- Processing is necessary for reasons of public interest.

The GDPR definition of biometric data does not expressly mention voiceprints. However, where a voiceprint is used as a unique identifier of a physical person it is hard to imagine that it would fall outside the ambit of Article 9. This interpretation is further supported by the express reference to 'dactyloscopic data' within the GDPR definition, and the obvious analogy between fingerprints and voiceprints.

GDPR Article 17 provides a data subject with a limited right (the so-called 'right to erasure') to have personal data held by a data controller erased. The Article is poorly drafted and creates considerable complexity. At its heart a data subject has the right to secure erasure of personal data concerning him/her without undue delay, provided that a number of conditions in relation to the data are met. Amongst these is where the data subject's consent (required under Article 9(2)(a)) has been withdrawn and where there is 'no other legal ground for the processing'. There are yet further, overarching, exceptions in relation to such things as the exercise of the right of freedom of expression, the public interest or official authority vested in the data controller, public interest in the area of public health, archiving purposes in the public interest or for scientific or historical research, or for the establishment, exercise, or defence of legal claims.

So although a voiceprint is unarguably a 'special category' of personal data under the GDPR, the rights that are provided, particularly as to the erasure of that voiceprint, are

more limited than would be the case if the person from whom the voiceprint has been taken had a form of proper 'ownership' over the voiceprint.

5.4. The 'uncaptured' voice

We now move further along our spectrum of tangibility away from recordings of voices or voiceprints created from recordings of voices, to the legal protection of an 'uncaptured' voice, where the sought protection is not over any particular recording but of a voice 'in general'.

Our first consideration is whether such an uncaptured voice can be protected as a registered trade mark.

5.4.1. Registered trade mark protection

We saw when we were looking at protection of a 'captured' voice that it is possible that an applicant seeking to register a distinctive voice saying a distinctive phrase in a distinctive manner could achieve a successful trade mark application. Submitting an audiofile of the voice saying that phrase in that manner would likely meet the 'clear and precise subject matter' test set out under the new EU trade mark law. We saw that the problem that such an application would encounter is that the voice, phrase and manner would have to be capable of distinguishing the goods or services of one undertaking from those of other undertakings.

However, what is clear from Sieckmann and following decisions (as discussed above) is that any attempt to obtain trade mark protection for a distinctive voice alone, i.e. not captured in any format, will fail. The new EU trade mark law will almost certainly not change that position. How does one go about representing a distinctive voice in the register 'in a manner which enables the registrar and other competent authorities and the public to determine the clear and precise subject matter of the protection afforded to the proprietor'? Merely stating 'the voice of Sir Sean Connery' in your application will clearly fail. What audio files would you submit? Any audio file you submit will merely be a single example of Sir Sean's voice – no single example alone would capture it sufficiently to meet the 'clear and precise subject matter test'. So, if a registered trade mark application will fail we need to go on to ask whether there is any way that the goodwill associated with a distinctive voice such as Sir Sean's can be protected.

5.4.2. The law of passing-off

The law of passing-off evolved in England and Wales, and some other common law jurisdictions, during the course of the nineteenth and twentieth centuries in parallel with registered trade mark law. Although there are obvious similarities and concerns between the two, being a common law (rather than a statutory) tort the scope of passing-off is significantly wider and its application can be very much more flexible. Passing-off is based upon the idea that it is wrong to mislead consumers into thinking that your product is actually somebody else's – essentially 'free-loading' off another's hard-earned goodwill. The requirements for bringing an action for passing-off in England and Wales were established by the UK House of Lords in Reckitt & Colman Products Ltd. v Borden Inc. [1990] 1 WLR 491. This case concerned Reckitt's wellknown lemon-shaped container for Jif lemon juice and Borden's decision to sell their

competing lemon juice in a similar-looking lemon-shaped container. The court held that to succeed with an action the claimant must establish:

i) goodwill or reputation in the goods or services by association with a particular 'get up' which would be recognised by the public as distinctive of the claimant's goods and services;

ii) there has been a misrepresentation which is likely to lead the public to believe the goods or services offered are those of the claimants; and

iii) that as a result damage has been (or will be) caused.

Accordingly, the law of passing-off will protect usual types of trade marks of the kind described above. Crucially, however, the protection is not restricted by the types of strict 'clear and precise subject matter' requirements that we saw for registered trade marks, so it will also protect other forms of marketing which support the use of standard trade marks, such as the way goods are packaged (so-called 'trade-dress') and presented, and such things as staff uniforms, trade stands, and shop designs. It is particularly effective where a number of the different elements of the marketing image of a brand have been used in conjunction by the passer-off. There is not then in theory any reason why the use of a distinctive voice as part of a marketing campaign, perhaps as a voice-over, could not be protected under the law of passing-off. However, there are significant potential problems. First of all, the voice would have to become very strongly associated with a product such that when consumers heard it (perhaps in certain contexts) they would bring that product to mind.

Let us imagine that Sir Sean Connery has long been used as a voice-over artist for advertisements for a range of Edinburgh-made sausages. It is certainly possible that many years of repeated use of Sir Sean's voice in relation to the sausages could build the required link in the mind of putative sausage purchasers. However, the linkage (and potential protection) would likely be very narrow. Would customers connect Sir Sean's voice with any other meat products? They might, but that would be a question of fact, and obviously the further one strayed from sausages – say, to pork pies – the connection would become more tenuous. Let us now imagine that a competitor sausage seller (perhaps a well-known German-headquartered discount retailer) uses a voice-mimic to imitate Sir Sean's voice in relation to a radio advertisement for sausages. Even if the mimicry were effectively perfect, would that constitute passingoff? It is likely that consumers would bring to mind the original 'Edinburgh' sausage, but would they believe that the German sausage was the Edinburgh sausage? It is unlikely they would conclude they were identical; the advert would clearly mention the name of the German retailer, and that would very likely to put the public on notice that the products had a different source. But might the public think that in some way the two were connected – perhaps that the Edinburgh sausage-makers had made the sausages for the Germans?

This was the reasoning behind the decision in the famous English High Court 'Penguin -v- Puffin' passing-off case (United Biscuits (UK) Ltd v Asda Stores Ltd [1997] RPC 513). The court held that although the public did not think that Asda's 'Puffin' chocolate biscuit was the 'Penguin' chocolate biscuit, they would have thought that there was a close connection; otherwise, Asda's activities would not have been allowed. It is fair to say that the public, in the UK at least, have become less naïve and are well aware of parody and 'cheeky' lookalikes (see Moroccanoil Israel Ltd v Aldi Stores Ltd [2014]

EWHC 1686 (IPEC)). If the mimicry of Sir Sean's voice was less than perfect, the public would be even less likely to be deceived.

Our imagined passing-off case assumed a clear link between a voice (Sir Sean) and a product (the sausage). However, the goodwill we are looking at here is the goodwill of the Edinburgh sausage maker. What, then, of goodwill in Sir Sean's voice itself? Celebrities of all sorts engage in endorsements of third-party products, often making more from their endorsement than the original cause of their celebrity. At the extreme end of endorsement, celebrities can also put their names to products, whether it be a George Foreman grill or a Beyoncé fragrance.

In the English Court of Appeal case of Irvine v Talksport Ltd ([2003] EWCA Civ 423); [2003] 2 All ER 881; [2003] EMLR 538) the radio station Talksport had mocked-up a photograph of racing driver Eddie Irvine holding a radio such that it looked like Irvine was listening to Talksport. Irvine claimed that it implied that he was endorsing the radio station. The court agreed with Irvine that although he was not a traditional trader as such, he did have commercial goodwill arising from his celebrity and his ability to use that celebrity to endorse products. As such, Talksport's action did constitute passing-off: the public would be deceived into thinking that Irvine was endorsing the station, which would damage Irvine's goodwill. If an unauthorised image of a celebrity can constitute passing-off of the celebrity's commercial goodwill, then there seems to be no reason, in principle, that an authorised voice-over by a voice mimic could constitute passing-off. There are, however, some important caveats.

First, the unauthorised mimicry would need to be so good as to mislead the vast majority of listeners into believing that the voiceover really was being provided by (say) Sir Sean. Something which was a parody or a 'cheeky' sound-alike, which (though it might bring Sir Sean to mind) would not lead the listener to believe that this was a genuine endorsement by Sir Sean, would be unlikely to found to be passing-off. It is of course hard to divorce the voice from what is being said. Messages that explicitly reveal the 'cheeky' nature of the sound-alike will in many cases take the voiceover even further away from being passing-off.

An interesting (unlitigated) example on recent British radio and television was the 2010 campaign for MoreThan insurance which used a relatively convincing soundalike of the American celebrity Morgan Freeman. The impersonator identified himself in the advertisements as 'MoreThan Freeman'. Would listeners think they were hearing a sound-alike or, notwithstanding that the voice on the advertisement was described as 'MoreThan Freeman', that it was still Morgan Freeman doing the voice work? In fact it was not Morgan Freeman who did the work – it was the US comedian and impressionist Josh Robert Thompson – but Freeman's consent was given, so in a sense this could be said to be a form of celebrity endorsement at (arguably) one remove.

This brings us to the second caveat: for a celebrity voice alone to carry endorsement goodwill it would in most cases likely need to be very distinctive indeed, such that in the absence any other clues to identity the listener would be able to put a name to the voice. Non-distinctive, unidentifiable celebrity voices would not, of themselves, carry endorsement goodwill without further prompts as to identity in the script (though in any event without a prompt they would likely not be used). It is important to note that the existence of a celebrity voice should not create any rights over generic accents or indeed over other individuals' use of their own voice. Though many will consider Sir Sean Connery's voice to be distinctive, elements of his voice will derive from his growing

up in the Fountainbridge district of Edinburgh with Irish grandparents, and the existence of Sir Sean's celebrity should not prevent any very similar-sounding native of Edinburgh from using his voice in a commercial context, unless there is some deliberate attempt to mislead consumers.

The third caveat is that passing-off requires the existence of commercial goodwill, and misrepresentation within a commercial or endorsement context does not create any right over a celebrity voice which is outwith this commercial sphere. It should not serve as a means for shutting down mimicry done for artistic or entertainment purposes. Indeed, in those cases the context is such that the audience knows that the mimic is not the celebrity.

5.4.3. 'Right in Publicity' in the United States

As can be seen in the test set out by the court in Reckitt & Colman, the law of passing-off in England and Wales is firmly based on the idea of preventing harm to the claimant's commercial goodwill. In Irvine the required harm was found to encompass damage to a slightly different sub-species of commercial goodwill, that which arises from a celebrity's endorsement activities. The protection of celebrity endorsement in the United States has taken a different evolutionary path. The principle which seems to underlie the approach taken in the US is to prevent the undue enrichment of the person using a celebrity's right, regardless of whether that use damages that celebrity's 'goodwill' (Dogan 2014).[9]

In Haelan Laboratories, Inc v Topps Chewing Gum Inc. (202 F.2d 866 (2d Cir. 1953) Haelan gained exclusive contracts for the use of baseball players' photographs in connection with the sale of Haelan's chewing gum. Haelan's competitor Topps induced the players to enter into new contracts for the use of their images by Topps. The celebrity's goodwill was not damaged, but Topps would become unduly enriched at the expense of Haelan, who had exclusivity over the images for sales of chewing gum. The United Second Circuit Court of Appeals held that there was a legal right in the value of a celebrity image which could be sold from one party to another.

Dogan (2014:18) asserts that 'Given the magnitude of its impact, Haelan was a remarkably terse decision – skimpy in its discussion of precedent, short on normative rationale, and utterly lacking in an examination of potential consequences.' Notwithstanding that, the right in Haelan has stood the test of time and has been developed by subsequent case law.

We saw in relation to the captured voice (section 5.2) that the court in Midler v Ford did not think that copyright could exist in a voice alone as this would not meet the fixation requirements for US copyright law, The court did, however, find that Midler's 'rights in publicity', which could as equally sit with her voice as with her photographic

---

[9] It is worth noting that many (including many civil law) jurisdictions achieve similar ends to the law of passing-off through doctrines of 'unfair competition' which are based upon the concept of unfair enrichment of the rogue trader, rather than the concept of harm to the claimant's goodwill.

image, had been infringed by Ford's actions. In the case of Waits v Frito-Lay, Inc. (978 F.2d 1093 (9th Cir. 1992) the advertising agency for acting for the snack-maker Frito Lay used a voice mimic to reproduce Tom Waits' distinctive gravelly voice in a song inspired by Waits' 1976 song Step Right Up for a radio commercial for Doritos snacks. Unsurprisingly, the court found this to be an infringement of the Midler right (Cecchin 1993).

Clearly, for a right in personality to exist in a voice alone, the voice will need to be very distinctive. However, we might define what this label might mean in phonetic and perceptual terms. Our question here is: how close does the mimicry have to be for there to be an infringement? Under our discussion of the English right of passing-off we suggested that a successful action would require more than the mimicked celebrity voice merely 'bringing to mind' the celebrity. Genuine confusion would be required. The position with the US 'right in celebrity' appears to different, reflecting the fact that the cause of action is not based on harm to the celebrity claimant, but on unfair gain by the defendant.

Dogan (2014:24) further states that 'even advertisements that parody, comment upon, or transform the celebrity image consistently run afoul of the celebrity's publicity rights', as was seen in White v Samsung Elecs. Am. Inc. (971 F.2nd 1395 9th Cir. 1992). Here, Samsung ran commercials in which a golden robot wearing a gown, jewellery, and a blonde wig, stood next to a 'Wheel of Fortune'-type letter board which clearly brought to the mind of the public the image of Vanna White, the very well-known host of the TV show Wheel of Fortune. The court held that using the likeness of a celebrity without consent was an infringement of the celebrity's right in personality, and that the right in publicity can extend to anything which invokes the celebrity's personality (likeness). However, there is a limit. Where the use of an image is part of an artistic work with a significant creative component, the defendant can rely on a freedom of artistic speech defence under the First Amendment to the U.S. Constitution (ETW Corporation v Jireh Publishing, Inc. 332 F.3d 915, 918 (6th Cir. 2003).

On the basis of the reasoning in Midler and Waits there seems to be no reason why the invocation principle in White would not apply to a distinctive voice.

Despite the differences in approach between the English law of endorsement passingoff under Irvine and the US right in publicity, there is one key similarity. Both only apply to a situation where a celebrity has some degree of recognition in the mind of the public. Although in neither situation is there a clear way of determining how much celebrity is required to sustain an action, it seems likely that recognition within a group of relevant consumers would be sufficient. Where the voice belongs to someone who cannot demonstrate any degree of 'celebrity' then neither law will provide any degree of protection. We will now look to the law of France which has (to an extent) developed a form of protection based on a person's inherent human rights.

5.4.4. The 'right of image' and the 'sonorous image' under French law

In 1969, Henri Charrière published a successful autobiographical novel, Papillon, about his imprisonment at and subsequent escape from a severe penal colony in French Guiana. In 1970, investigative journalist Gérard de Villiers published Papillon Épinglé ('Butterfly Pinned'), which challenged the accuracy of Charrière's account. Some of Butterfly Pinned was based on court documents concerning Charrière's original

sentencing to French Guiana. Logeais & Schroeder (1998) report that in the Papillon case (TGI Paris ord., réf., Feb. 27, 1970) the court held that the publication of Butterfly Pinned did not amount to an invasion of Charrière's privacy. In contrast, however, the court held that the presence of a photograph of Charrière on the cover of the book, without his consent, infringed upon Charrière's 'right of image', and suggested that Charrière had given endorsement to de Villiers' book (which, of course, he had not).

On first examination this looks something like 'endorsement passing off' in the case in Irvine, or the 'right of publicity' established in the United States in Haelan, i.e. something which requires a degree of celebrity. For Logeais & Schroeder (1998): 'most of the decisions recognizing a right on the image [or the right to profit on the image] in France involve celebrities, whether or not well-known, for whom their image is an essential feature of their professional career, including performing artists, athletes, and models. In these cases, courts have acknowledged a right to authorize and control such use. Inexplicably, however, they have failed to award adequate (or substantial) damages.' (1998:517).

However, the French right to control an image appears to arise out Article 9 of the Civil Code of France (inserted by Act of Parliament of 17 July 1970), which provides that 'everyone has the right to respect for his or her private life'. Given that Article 9 refers to 'everyone', is this 'right of image' more akin to a right of privacy which would protect those who have no celebrity?

Logeais & Schroeder (1998: 517) report that the right to protect an image in France 'is characterized as a personality right rather than as a property right, despite some courts' contrary statements'. Lefranc (2014) further comments that the exact nature of the right is unclear, in that there has been no one decision which lays a

> … foundation for an exclusive right in the market value of celebrity identity. Instead there are innumerable decisions which in fact apply such a right, without ever explaining why, because French judges are very reluctant to articulate principles and lay down rules. (2014:39)

Caron (2006) suggests that French law respects two rights in an image: one a patrimonial right over image which 'only makes sense for individuals whose image has a commercial value, such as celebrities, as well as individuals who are temporarily under media scrutiny', and the other a traditional extra-patrimonial 'moral' right of image, allowing all individuals, irrespective of celebrity, to oppose alterations to their image.

How does this 'right of image' (whatever its exact nature and effect) impact on the protection of a voice? Lepage (2005:2) reports that in the 2004 decision of the Tribunal de Grande Instance de Paris, (TGI Paris, 3e ch., 1re sect., 27 sept. 2004) the court held that:

> …by application of Article 9 of the Civil Code, everyone has the right to respect for one's private life; this right also covers the right of image, of name and of voice. [our translation]

Lepage states further that:

> Certain decisions recognise the voice as a personal attribute (TGI Paris, 3e ch., 1re sect., 27 sept. 2004: préc.) whilst others suggest that there is a right of voice, detached from the right to respect for one's private life. This is when the analogy with the image is introduced – in fact, it is revealed by the expression 'sonorous image' [image sonore] which is sometimes used when speaking about the voice. [our translation]

Tafforeau (2007) reports on the judgment of November 22, 2006 by the Tribunal de Grande Instance in Paris, 3e, in relation to the Cannes Film Festival award-winning documentary film Être et Avoir ('Be and Have'). The film followed a teacher and his pupils during their lessons and school trips, as well as parents helping their children with homework. Following the film's success at Cannes, the director released a DVD on which there was a bonus feature about the making of the film. The parents claimed that there were problems with the original consents given for the main film, and that no permission had been given for the use of the images and voices of the parents and their children in the bonus feature. The court held that 'patrimonial' image rights would only apply when the image of the person in question is well known. When the original contracts were signed none of the students or parents was widely familiar to members of the general public; it was the success of the film that made them famous. Therefore, the court rejected the claimants' complaints with regard to the main film, but upheld their claim in respect of the DVD bonus feature. For Tafforeau, recognising the existence of a patrimonial right in the claimants' image arose out of their 'fame'.

Clearly, however, the parents and children in Être et Avoir had relatively little fame when we compare it to that of Bette Midler or Tom Waits in the US cases discussed earlier. Neither could we really say that the rights of the Être et Avoir participants were based upon the distinctiveness of their voices. It seems improbable that the public at large, unless perhaps they had watched the film multiple times, would have been able reliably to pick out the voices of the children or parents in a 'voice parade' line-up, for example (see further McGorrery & McMahon 2017; Robson 2017).

So one might argue that French law, although creating a 'right of sonorous image' which is supposedly based upon a fundamental right to individual privacy, has reached a confused position where it has shied away from granting protection to those that have no degree of fame, but which will grant protection on the basis of very little fame and, it would appear, with no requirement for distinctiveness.

5.4.5. Fraud and fraudulent misrepresentation

We discussed above (section 4) the situation where voice samples are used for voice cloning. We concluded that the activity would be fraud where such voice cloning was intended to dishonestly make a false representation and intended by making the representation (a) to benefit the voice cloner or another, or (b) to cause loss to another, or (c) to expose another to a risk of loss. The situation is of course no different where the false representation does not involve recorded voice samples but is merely someone seeking to accurately mimic a voice 'live'. Again, in determining how much vocal similarity is required, our concern is about the intention of the fraudster rather than

similarity as such. We also noted that a voiceprint (or digital spoofing) could give rise to a civil claim under the tort of fraudulent misrepresentation, and as before the situation is no different where the false representation does not involve recorded voice samples but is the action of someone using his or her own voice to imitate the vocal patterns of another individual.

6. Summary and conclusions

Table 1 summarises the ways in which a voice can be protected under the laws we have examined. Some of these protections, such as fraud, fraudulent misrepresentation, and data protection, are not strictly rights of ownership, in that they do not create a protection that can be 'traded' like a commodity. The others, by contrast, are tradeable commodities.

The table reveals some patterns. In relation to the protections provided for misuse of voiceprints the tests for the existence of the right are, as one might expect, objective, and they do not require any subjective judgment as to voice distinctiveness. The protection of the uncaptured voice reveals the broadest range of types of protection. Predictably, the protection of the uncaptured voice (other than the confusing 'right of image' in France and protection provided under fraud, which is intention-based) requires some degree of subjective determination of voice distinctiveness. Clearly, some aspects of the right of image in France will require a subjective determination of voice distinctiveness, but that may not be consistent in all applications of the right.

The protection of the captured voice through Registered Trade Marks mirrors aspects of the English right in passing-off and indeed the US Right of Celebrity. Both of the last two of these serve, in slightly different ways, to protect celebrity goodwill. Trade marks can do the same, but face a limitation in that trade marks are required to meet a 'clear and precise subject matter' test.

Tangibility	'Right'	Objective test	Subjective test of voice distinctiveness
Captured Voice (Recording)	Copyright	Fixation creates right, not distinctiveness.	

	Registered Trade Mark		Distinctiveness must be capable of distinguishing the goods or services of one undertaking from those of other undertakings.
Voiceprint and computer-based models	Copyright	De minimis rule / 'expression of the author's intellectual creation.' Likely to be independent of distinctiveness.	
	Fraud	The intention of the fraudster is key; not reliant on distinctiveness.	
	Fraudulent misrepresentation	Misrepresentation needs to induce a party into entering into a contract.	
	Data protection	Applies to all biometric data.	
Uncaptured Voice (Live mimicry or digital spoofing)	Passing-off		Requires mimicry to be so good as to mislead the vast majority of listeners as to origin of voice. Not sufficient to merely 'bring celebrity to mind'.
	Right of Publicity (US)		Sufficient merely to bring celebrity to mind.
	Right of Image (France)		May not be based on distinctiveness at all.
	Fraud	The intention of the fraudster is key; not reliant on distinctiveness.	
	Fraudulent misrepresentation		Misrepresentation needs to induce a party into entering into a contract; may require element of voice distinctiveness.

Table 1. Ways in which a voice can be protected under the law, ranked in terms of their 'tangibility'. Commercially tradable rights are in bold and underlined. (England & Wales unless otherwise noted).

It will be apparent from the foregoing discussion that pinning down the human voice in a way that would meet a clarity and precision test, or other tests of a similar kind, is intrinsically very difficult. Defining the boundaries of an individual speaker's voice in a satisfactorily objective way is at the heart of the problem. It is one thing to capture

audio samples and to produce visual representations of them in the form of spectrograms. These are relatively tangible objects. But what of the abstract statistical models generated by contemporary speaker verification and comparison systems? These are no more 'the voice' of a speaker than a string of numbers representing image properties such as pixel hue, saturation, and lightness in digital photographs are 'the face' of the subject. And yet these speaker models have proven to be sufficiently good representations of talkers' speech that they are routinely used for a range of high-risk applications. As Sanjith reminds us, in the context of the successful livemimicry spoofing attempt by Dan and Joe Simmons, 'There is no "silver-bullet" solution on the fight against fraud; and sometimes it takes a couple of very ingenious pranksters to remind us never to let our guard down. […] But today, as far as voice biometric implementations go, millions of successes are taken for granted and failure has a million critics' (Sanjith 2017). When security breaches occur – on the occasions that they are publicised, at any rate – a loss of confidence in voice biometric technology on the part of the customer base is liable to follow. That damage can be hard to repair. But public anxieties can be allayed to at least some degree if consumers feel reassured that their rights are protected by legislation that has been regularly revised in reaction to, and anticipation of, attacks by fraudsters. It is imperative, we would submit, that clearer and more appropriate provisions be put in place to supersede what sometimes appears to be fudged repurposing of existing, often vague, legislation. We echo Mirra's (2018) warning against complacency, and his call for positive action:

> The law must adapt to reconcile the evolving possibilities of tomorrow instead of being entrenched in the antiquated shadow of yesterday. In order to prevent the voice recordings from being distorted by a few keystrokes on a laptop, the law must account for potential foul play facilitated by new technologies. As novel innovations in technology proliferate, opportunities for dishonesty multiply. Project [Adobe] VoCo and similar [voice cloning] technologies are coming, and it will demand change. (Mirra 2018:30)

We have in this paper only considered the situation in three jurisdictions: principally England and Wales (and where appropriate the EU where it is a source of English law), and particular applications of the law of the United States and France. The situation is likely to be significantly different in other countries, though to the best of our knowledge there are no legal codes anywhere that protect uncaptured voice in a more explicit way than do those of the US and France. Further comparisons which attempt to keep abreast of developments in line with societal and technological changes in the relevant states will form the basis for future research.

References

Akinyokun, N. & Teague, V. (2017). Security and privacy implications of NFCenabled contactless payment systems. Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES), Reggio Calabria, Italy, September 2017, Article 47. DOI: 10.1145/3098954.3103161.

Armerding, T. (2018). The 18 biggest data breaches of the 21st century. CSOonline, 20th December 2018. Online resource: <www.csoonline.com/article/2130877>, accessed 4th May 2019.

Belin, P., Fecteau, S. & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. Trends in Cognitive Sciences 8(3): 129-135.

Blair, O. (2017). Australian radio DJ says royal prank call to Duchess of Cambridge's hospital contributed to her divorce. The Independent, 27th April 2017. Online resource: <https://www.independent.co.uk/life-style/australian-radio-dj-mel-greig-bbc-royalprank-call-duchess-cambridge-hospital-nurse-jacintha-a7706186.html>, accessed 4th May 2019.

Bohm, A., George, E., Cyphers, B. & Lu, S. (2017). Privacy and liberty in an alwayson, always-listening world. Columbia Science and Technology Law Review 19: 145. Online resource: <http://www.stlr.org/download/volumes/volume19/Bohm.pdf>, accessed 4th May 2019.

Bolt, R.H., Cooper, F.S., David, E.E., Denes, P.B., Pickett, J.M. & Stevens, K.N. (1970). Speaker identification by speech spectrograms: A scientists' view of its reliability for legal purposes. Journal of the Acoustical Society of America 47(2): 597-612.

Caron, C. (2006). Un nouveau droit voisin est né: Le droit patrimonial sur l'image. Communication Commerce électronique, January 2006, no. 1. Online resource: < https://www.lexis360.fr/Document/droits_voisins_un_nouveau_droit_voisin_est_ne_le_droit_patrimonial_sur_limage/gJMzhyZhK6sytawyL6uTRy5EpSF_W-3tNPAhKjUCi1I1?data=c0luZGV4PTEmckNvdW50PTExNzcm&rndNum=1972115659&tsid=search4_>, accessed 4th May 2019.

Cecchin L.M. (1993). Waits v. Frito-Lay, Inc. and Tracy-Locke, Inc. 978 F.2d 1093 (9th Cir. 1992). DePaul Journal of Art, Technology and Intellectual Property Law 3(2): 88-90.

Chen, S., Ren, K., Piao, S., Wang, C., Wang, Q., Weng, J., Su, L. & Mohaisen, A. (2017). You can hear but you cannot steal: Defending against voice impersonation attacks on smartphones. Proceedings of the 37th Institute of Electrical and Electronics Engineers (IEEE) International Conference on Distributed Computing Systems, Atlanta, June 2017, pp. 183-195. Online resource: <https://cse.buffalo.edu/~lusu/papers/ICDCS2017Si.pdf>, accessed 4th May 2019.

Chouiter, L. & Annoni, J.-M. (2018). Glossolalia and aphasia: Related but different worlds. In Bogousslavsky, J. (ed.). Neurologic-Psychiatric Syndromes in Focus, Part II: From Psychiatry to Neurology. Basel: Karger, pp. 96-105.

Cowling, M. (2015). It's not just your TV listening in to your conversation. The Conversation, 10th February 2015. Online resource: < http://theconversation.com/its-not-just-your-tv-listening-in-to-your-conversation37409>, accessed 4th May 2019.

Crime Survey for England and Wales (2017). Crime in England and Wales: Year Ending December 2017. Online resource: <https://www.gov.uk/government/statistics/crime-in-england-and-wales-yearending-december-2017>, accessed 4th May 2019.

Crime Survey for England and Wales (2018). Crime in England and Wales: Year Ending December 2018. Online resource:

<https://www.ons.gov.uk/releases/crimeinenglandandwalesyearendingdecember2018>, accessed 4th May 2019.

Curia Press Release 106/03 re Shield Mark BV v Joost Kist h.o.d.n. MEMEX (Case C-283/01, November 2003). Online resource: <https://curia.europa.eu/en/actu/communiques/cp03/aff/cp03106en.htm>, accessed 4th May 2019

Davies, C. (2014). Jacintha Saldanha 'took blame' for Duchess of Cambridge prank call. The Guardian, 11th September 2014. Online resource: <https://www.theguardian.com/world/2014/sep/11/jacintha-saldanha-took-blameprank-call-duchess-cambridge-australian-djs-inquest>, accessed 4th May 2019.

Dogan, S.L. (2014). Haelan Laboratories v. Topps Chewing Gum: Publicity as a legal right. In Cooper Dreyfuss, R. & Ginsburg, J.C. (eds). Intellectual Property at the Edge: The Contested Contours of IP. Cambridge: Cambridge University Press, pp. 17-38.

Eckert, H. & Laver, J. (1994). Menschen und ihre Stimmen: Aspekte der Vokalen Kommunikation. Weinheim: Beltz PVU.

Feng, H., Fawaz, K. & Shin, K. (2017). Continuous authentication for voice assistants. Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, Snowbird, UT, October 2017, pp. 343-355.

Gold, E. & French, P. (2011). International practices in forensic speaker comparison. International Journal of Speech, Language and the Law 18(2): 293-307.

Greenslade, R. (2011). News of the World pays footballer £70,000 for libel. The Guardian, 19th June 2011. Online resource: <https://www.theguardian.com/media/greenslade/2011/jun/19/newsoftheworldmedialaw>, accessed 4th May 2019.

Harrington, J. (2006). An acoustic analysis of 'happY-tensing' in the Queen's Christmas broadcasts. Journal of Phonetics 34(4): 439-457.

Hashim, N.W., Wilkes, M., Salomon, R., Meggs. J. & France, D.J. (2017). Evaluation of voice acoustics as predictors of clinical depression scores. Journal of Voice 31(2): 256.e1-256.e6.

Hollien, H. (1990). The Acoustics of Crime: The New Science of Forensic Phonetics. New York: Springer.

HSBC UK (2019). Telephone banking: Keep on top of your money by phone. Online resource: <https://www.hsbc.co.uk/ways-to-bank/phone-banking/>, accessed 4th May 2019.

Ihalainen, J. (2018). That sounds good - do you have IP rights in your own voice? IP Iustitia, 31st January 2018. Online resource: <https://www.ipiustitia.com/2018/01/that-sounds-good-do-you-have-iprights.html>, accessed 4th May 2019.

Jones, R. (2018). Voice recognition: is it really as secure as it sounds? The Guardian, 22nd September 2018. Online resource: <https://www.theguardian.com/money/2018/sep/22/voice-recognition-is-it-reallyas-secure-as-it-sounds>, accessed 4th May 2019.

Kasuya, H. & Yoshida, H. (2017). Age-related changes in the human voice. In Makiyama, K. & Hirano, S. (eds.). Aging Voice. New York: Springer, pp. 27-36.

Kersta, L.G. (1962). Voiceprint identification. Nature 196: 1253-1257.

Kersta, L.G. (1966). Speaker recognition and identification by voiceprints. Connecticut Bar Journal 40(4): 586-593.

Keulen, S., Mariën, P., van Dun, K., Bastiaanse, R., Manto, M. & Verhoeven, J. (2017). The posterior fossa and Foreign Accent Syndrome: Report of two new cases and review of the literature. The Cerebellum 16(4): 772-785.

Kreiman, J. & Sidtis, D. (2011). Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception. Oxford: Wiley.

Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics. Chicago: University of Chicago Press.

Lavan, N., Burton, A.M., Scott, S.K. & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. Psychonomic Bulletin & Review 26(1): 90-102.

Laver, J. (2017). Linguistic phonetics: The sounds of languages. In Aronoff, M. & Rees-Miller, J. (eds.). The Handbook of Linguistics, 2nd edn. Oxford: Wiley, pp. 159-184.

Loakes, D. (2006). Variation in long-term fundamental frequency: Measurements From vocalic segments in twins' speech. Proceedings of the 11th Australian International Conference on Speech Science & Technology, Auckland, December 2006, pp. 205-210.

Lefranc, D. (2014). Do the French have their own "Haelan" case? The droit à l'image as an emerging intellectual property right. In Cooper Dreyfuss, R. & Ginsburg, J.C. (eds.). Intellectual Property at the Edge: The Contested Contours of IP. Cambridge: Cambridge University Press, pp. 39-56.

Lepage, A. (2005). Protection de la voix à travers celle de la vie privée. Communication Commerce électronique, May 2005, no. 5: 48. Online resource: < https://www.lexis360.fr/Document/voix_protection_de_la_voix_a_travers_celle_de_la_vie_privee_commentaire/DF08shty4yS0fq6Mz8EttXA4r2nKa2Mi6K7OYiC7EEU1?data=c0luZGV4PTEmckNvdW50PTIzMjAm&rndNum=1653368489&tsid=search5_>, accessed 22nd May 2019.

Logeais, P. and Schroeder, J.-P. (1998). The French right of image: an ambiguous concept protecting the human persona. Loyola of Los Angeles Entertainment Law Review 18(3): 511-542.

McClelland, E. (2011). Impersonation in forensic casework: Case of Tommy Sheridan. Paper presented at the annual conference of the International Association for Forensic Phonetics and Acoustics, Vienna, Austria, July 2011. Online resource: <https://www.kfs.oeaw.ac.at/publications/iafpa_abstracts/nr26_elizabethmcclellan d.pdf>, accessed 5th May 2019.

McGettigan, C. & Lavan, N. (2017). Human voices are unique – but our study shows we're not that good at recognising them. The Conversation, 16th June 2017. Online resource: <https://theconversation.com/human-voices-are-unique-but-our-studyshows-were-not-that-good-at-recognising-them-79520>, accessed 5th May 2019.

McGorrery, P.G. & McMahon, M. (2017). A fair 'hearing': Earwitness identifications and voice identification parades. International Journal of Evidence and Proof 21(3): 262-286.

McLeod, K. & DiCola, P. (2011). Creative License: The Law and Culture of Digital Sampling. Durham, NC: Duke University Press.

Mermelstein, P. (1976). Distance measures for speech recognition psychological and instrumental. In Chen, C.H. (ed.). Pattern Recognition and Artificial Intelligence. New York: Academic Press, pp. 374-388.

Miller, D. (2016). Voice Biometrics Census: Steady Growth of Global Enrollments. San Francisco: Opus Research, Inc. Online resource: <https://www.nuance.com/content/dam/nuance/en_au/collateral/enterprise/report/ar-opus-vb-census-en-us.pdf >, accessed 5th May 2019.

Mirra, N. (2018). Putting words in your mouth: The evidentiary impact of emerging voice editing software. Richmond Journal of Law and Technology 25(1): 1-30.

Morrison, G.S., Sahito, F.H., Jardine, G., Djokic, D., Clavet, S., Berghs, S. & Dorny, C.G. (2016). INTERPOL survey of the use of speaker identification by law enforcement agencies. Forensic Science International 263: 92-100.

Mukhopadhyay, D., Shirvanian, M. & Saxena, N. (2015). All your voices are belong to us: Stealing voices to fool humans and machines. In Pernul, G., Ryan, P.Y.A. & Weippl, E. (eds.). Computer Security - ESORICS 2015 (20th European Symposium on Research in Computer Security, Vienna, September 2015), Proceedings, Part II, pp. 599-621.

Nolan, F., McDougall, K. & Hudson, T. (2011). Some acoustic correlates of perceived (dis)similarity between same-accent voices. In Lee, W.-S. & Zee, E. (eds.). Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, August 2011, pp. 1506–1509.

Nuance Communications (2018). Nuance Forensics: Prosecute Criminals using their Voice. Burlington, MA: Nuance Communications. Online resource: <https://www.nuance.com/content/dam/nuance/en_us/collateral/enterprise/datasheet/ds-nuance-forensics-en-us.pdf>, accessed 5th May 2019.

Paulsen, F., Kimpel, M., Lockemann, U. & Tillmann, B. (2000). Effects of ageing on the insertion zones of the human vocal fold. Journal of Anatomy 196: 41-54.

Peachey, K. (2019). HMRC forced to delete five million voice files. BBC News, 3rd May 2019. Online resource: <https://www.bbc.co.uk/news/business-48150575>, accessed 6th May 2019.

Ping, W., Peng, K., Gibiansky, A., Arık, S.O., Kannan, A., Narang, S., Raiman, J. & Miller, J. (2018). Deep Voice 3: Scaling text-to-speech with convolutional sequence learning. Proceedings of the 6th International Conference on Learning Representations (ICLR), Vancouver, April-May 2018. Online resource: <https://arxiv.org/pdf/1710.07654.pdf>, accessed 5th May 2019.

Potter, R.K., Kopp, G.A. & Green, H.C. (1947). Visible Speech. New York: Van Nostrand.

Rhodes, R. (2017) Aging effects on voice features used in forensic speaker comparison. International Journal of Speech, Language and the Law 24(2): 177199.

Roberts, L. (2011). Acoustic effects of authentic and acted distress on fundametal frequency and vowel quality. Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong, August 2011, pp. 1694-1697.

Robson, J. (2017). A fair hearing? The use of voice identification parades in criminal investigations in England and Wales. Criminal Law Review 1: 36-50.

San Segundo, E. & Mompean, J.A. (2017). A simplified Vocal Profile Analysis protocol for the assessment of voice quality and speaker similarity. Journal of Voice 31(5): 644.e11-644.e27.

Sanjith, R. (2017). BBC prank of HSBC voice ID system overlooks millions of successful authentications. Opusresearch.net, 22nd May 2017. Online resource: <https://opusresearch.net/wordpress/2017/05/22/bbc-prank-of-hsbc-voice-idsystem-overlooks-millions-of-successful-authentications/>, accessed 5th May 2019. Scherer, K. (1995). Expression of emotion in voice and music. Journal of Voice 9(3): 235-248.

Sewell, A. (2014). How copyright affected the musical style and critical reception of sample-based hip-hop. Journal of Popular Music Studies 26(2-3): 295-320.

Shirvanian, M., Saxena, N. & Mukhopadhyay, D. (2018). Short voice imitation manin-the-middle attacks on Crypto Phones: Defeating humans and machines. Journal of Computer Security 26(3): 311-333.

Sidtis, D. & Kreiman, J. (2012). In the beginning was the familiar voice: Personally familiar voices in the evolutionary and contemporary biology of communication. Integrative Psychological and Behavioral Science 46(2): 146-159.

Simmons, D. (2017). BBC fools HSBC voice recognition security system. BBC News, 19th May 2017. Online resource: <https://www.bbc.co.uk/news/technology39965545>, accessed 5th May 2019.

Singh, R., Jiménez, A. & Øland, A. (2017). Voice disguise by mimicry: Deriving statistical articulometric evidence to evaluate claimed impersonation. IET Biometrics 6(4): 282-289.

Smith, K.M. & Caplan, D.N. (2018). Communication impairment in Parkinson's disease: Impact of motor and cognitive symptoms on speech and language. Brain and Language 185: 38-46.

Sundberg, J. (1987). The Science of the Singing Voice. DeKalb, IL: Northern Illinois University Press.

Tafforeau, P. (2007) 'L'être et l'avoir ou la patrimonialisation de l'image des personnes' Communication Commerce électronique, Mai 2007, no. 5. Online resource: <https://www.lexis360.fr/Document/droit_a_limage_letre_et_lavoir_ou_la_patrimonialisation_de_limage_des_personnes_/IXNdyFojNyAsFUqe8SCteDD2fDiwGqmXVsgB2pHUghs1?data=c0luZGV4PTEmckNvdW50PTExMyY=&rndNum=715439695&tsid=search6_>, accessed 22nd May 2019.

Torre, P. & Barlow, A.J. (2009). Age-related changes in acoustic characteristics of adult speech. Journal of Communication Disorders 42(5): 324-333.

Tosi, O., Oyer, H.J., Lashbrook, W., Pedney, C., Nichol, J. & Nash, W. (1972). Experiment on voice identification. Journal of the Acoustical Society of America 51(6): 2030-2043.

Veaux, C., Yamagishi, J. & King, S. (2015). A comparison of manual and automatic voice repair for individual with vocal disabilities. Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT 2015), Dresden, September 2015, pp. 130-133.

Watt, D. & Brown, G. (2019). The forensic application of Automatic Speaker Recognition technology. To appear in Coulthard, M. & Johnson, A. (eds.). The Routledge Handbook of Forensic Linguistics, 2nd edn. London: Routledge.

Wenndt, S.J. (2016). Human recognition of familiar voices. Journal of the Acoustical Society of America 140: 1172-1183.

Wester, M., Wu, Z. & Yamagishi, J. (2015). Human vs. machine spoofing detection on wideband and narrowband data. Proceedings of the 16th Annual Conference of the

International Speech Communication Association (Interspeech), Dresden, September 2015, pp. 2047-2051.
Wu, Z., Evans, N., Kinnunen, T., Yamagishi, J., Alegre, F. & Li, H. (2015). Spoofing and countermeasures for speaker verification: A survey. Speech Communication 66: 130-153.
Yorkshire Post (2009). Gambling addict murdered dad and left his body in bin bags. Yorkshire Post, 20th February 2009. Online resource: <https://www.yorkshirepost.co.uk/news/latest-news/gambling-addict-murdereddad-and-left-his-body-in-bin-bags-1-2335486>, accessed 5th May 2019.
Zetterholm, E. (2007). Detection of speaker characteristics using voice imitation, In Müller, C. (ed.). Speaker Classification II: Selected Projects. Lecture Notes in Computer Science 4441. Berlin: Springer, pp. 192-205.