Rivista Trimestrale di SCIENZA dell'AMMINISTRAZIONE
Studi di teoria e ricerca sociale
4/2020

Le orecchie della smart city. Riconoscimento vocale e ascolto operazionale nella "città senziente"

Domenico Napolitano
PhD student, Università Suor Orsola Benincasa di Napoli

Riassunto

Questo contributo intende investigare da un punto di vista teorico-critico una delle facoltà strategicamente decisive per l'amministrazione della città del futuro, benché tra le meno studiate: la facoltà di ascoltare. Laddove la metafora dell'ascolto è stata impiegata in riferimento alla smart city come luogo della democrazia e della partecipazione civica, un'analisi tecnica dei modi in cui la città ascolta e origlia permette di individuare, incorporati negli apparati tecnologici, punti di contraddizione con quelle rappresentazioni. Partendo dai casi studio di dispositivi di monitoraggio acustico introdotti sperimentalmente a Santander (EAR-IT) e New York (SONYC), l'articolo ricostruisce l'idea di ascolto soggiacente a quei sistemi, al fine di decostruire narrazioni stereotipate sull'ascolto macchinico e sull'intelligenza artificiale. In primo luogo prende in considerazione il concetto di "ascolto operazionale" come pratica basata sul machine learning che ridefinisce l'ascolto umano nella sua interazione con l'ascolto macchinico. In secondo luogo analizza il funzionamento della sorveglianza acustica attraverso il riconoscimento sonoro, l'identificazione del parlante e l'individuazione di parole chiave. Da queste analisi l'articolo individua nella sorveglianza acustica, invisibile e ambientale, un esempio di sorveglianza post-panottica, ovvero non orientata al disciplinamento dei soggetti, ma all'estrazione di quanti più dati possibile per alimentare sistemi algoritmici di previsione. Essa mette dunque in collegamento diretto le modalità statistiche della governance con una nuova configurazione securitaria della società, orientata non tanto all'imputazione e alla repressione, quanto alla previsione, alla prevenzione e alla valutazione. Ciò incide profondamente sul rapporto tra soggettività e privacy, in quanto la prima è sempre più definita dalla sua capacità di fornire dati, diventando un medium per il funzionamento autonomo delle macchine.

Parole chiave: ascolto operazionale, audio monitoring, riconoscimento vocale, sorveglianza ubiqua, post-panopticon, governamentalità algoritmica

Abstract. *The Ears of the Smart City. Speech Recognition and Operational Listening in the "Sentient City"*

This contribution wants to investigate, from a theoretical-critical point of view, one of the strategically decisive faculties for the administration of the city of the future, although among the least studied: the faculty of listening. Where the metaphor of listening has been used in reference to the smart city as a place of democracy and civic participation, a technical analysis of the ways in which the city listens and eavesdrops allows us to identify points of contradiction with those representations that have been incorporated into the technological apparatus. By focusing on the case studies of acoustic monitoring devices introduced experimentally in Santander (EAR-IT) and New York (SONYC), the paper reconstructs the idea of listening underlying those systems in order to deconstruct stereotyped narrations about machine listening and artificial intelligence. In the first place, it takes in consideration the concept of "operational listening", as a practice of machine learning that redefines human listening in its interaction with machine listening. Then it analyzes the functioning of acoustic surveillance through sound recognition, identification of the speaker and keyword spotting. Starting from those analysis, the paper individuates in the environmental and invisible features of acoustic surveillance the paradigmatic traits of a post-panoptic surveillance, that is a surveillance not oriented towards the disciplining of subjects, but to the extraction of as much data as possible to feed algorithmic prediction systems. It therefore connects the statistical methods of governance with a new securitarian configuration of the society, oriented not so much to imputation and repression, as to prediction, prevention and evaluation. This affects deeply the relation between subjectivity and privacy, since the former is ever more defined by the data it can provide, so becoming ever more a medium for machines autonomous functioning.

Keywords: operational listening, audio monitoring, speech recognition, ubiquitous surveillance, post-panopticon, algorithmic governmentality

DOI: 10.32049/RTSA.2020.4.04

1. Introduzione

L'ultimo decennio ha visto affermarsi in contemporanea l'idea di smart city e gli studi critici che ne mettono in discussione, a seconda dei casi, la validità, la funzionalità, il modello economico, la retorica. Per alcuni, la smart city fa principalmente riferimento a un

uso assennato e sostenibile delle risorse, a un'ottimizzazione tecnologica dei servizi in tempo reale che rende la vita cittadina più sicura, efficiente, affascinante, mentre per altri essa è la materializzazione di un modello di economia neoliberista che riproduce e inasprisce l'ineguaglianza sociale. Se la pubblicistica accoglie con toni quasi sempre entusiastici l'avvento delle tecnologie smart nell'ambito cittadino, la letteratura scientifica ha mosso una serie di critiche al nuovo modello. Rob Kitchin (2014) mette in evidenza alcuni punti deboli della smart city, quali la tendenza alla tecnocrazia, la corporatizzazione della governance da parte dei colossi della tecnologia che, attraverso i loro servizi, rendono "smart" le città, la hackerabilità di quei sistemi come nuovo pericolo per la sicurezza, la tendenza al controllo ubiquo dei cittadini come condizione per il funzionamento dei servizi informatici. Eugeny Morozov (2018) sottolinea il carattere essenzialmente neoliberista della smart city e la condizione di crescente dipendenza di questa da servizi informatici che, mentre effettuano le loro operazioni, valutano anche la performance della città stessa, asservendola a una logica del ranking col ricatto della valutazione per l'accesso al credito. Iaconesi e Persico (2015) mettono in evidenza la dimensione esperienziale del soggetto che vive nella smart city, i cui movimenti, la cui rete sociale, le cui preferenze, sono orientati dalla mediazione di sistemi a base di dati non privi di bias tanto tecnici quanto ideologici. Anthony Townsend (2014) spiega le connessioni tra smart city e big data prodotti dagli smartphone, riconfigurando la posizione del cittadino da attore e destinatario delle politiche locali a medium tra amministrazioni locali e servizi informatici transnazionali. Mark Andrejevic (2019) sottolinea, inoltre, come la smart city sia la destinazione delle tecnologie di apprendimento automatico non solo a livello economico-politico, ma anche epistemologico, in quanto luogo in cui si materializza un'idea di sapere sempre più legata alle propensioni e alle possibilità macchiniche, quali il riconoscimento di *pattern*, la misurazione puntuale degli eventi e l'intervento automatico.

Appoggiandosi al quadro critico appena descritto, il presente studio intende indagare un aspetto particolare della smart city, ma che si rivela emblematico per la definizione delle condizioni di possibilità tanto tecniche quanto epistemologiche della nuova città intelligente e delle tecnologie che la presiedono. Si tratta dell'*ascolto*, una delle facoltà strategicamente

decisive per l'amministrazione della città del futuro, che mette in campo sia immaginari sociali che precise logiche comunicative interpersonali, uomo-macchina e macchina-macchina.

A un livello metaforico, l'ascolto è inteso come la capacità della smart city di comprendere la parola come luogo privilegiato d'informazione, e di assurgere, attraverso le tecnologie che mettono in contatto cittadini e amministrazione, a emblema della democrazia e della partecipazione civica (Simonofski *et al.*, 2019; Unione Europea, 2011). Approcci più pragmatici, invece, insistendo sull'analogia con l'udito umano, considerano l'ascolto come uno strumento di sorveglianza, spionaggio e invasione del privato, ma anche di monitoraggio preventivo che può incrementare il senso di sicurezza. Questo studio intende distaccarsi da entrambe le visioni per problematizzare le specificità dell'*ascolto macchinico* e i rapporti di continuità e discontinuità di questo con i paradigmi teorico-critici sull'ascolto, le tecnologie e la società.

Attraverso un'analisi dei dispositivi tecnici di ascolto macchinico e i casi specifici del loro impiego sia nell'ambito domestico, come gli ormai noti assistenti virtuali vocali Amazon Alexa o Google Home, sia nell'ambito cittadino, definiti all'occorrenza *acoustic event detectors*, *acoustic sensing monitors* o sensori di *audio monitoring*, e attraverso il confronto di questi con gli ordini discorsivi che li definiscono, questo studio intende mettere in luce le condizioni epistemologiche aperte dall'ascolto macchinico e le conseguenze socio-culturali che questo produce. I concetti di comunicazione, sorveglianza, privacy, verranno dunque riletti alla luce dei saperi incorporati negli apparati tecnologici della città "in ascolto", mettendo in evidenza punti di contraddizione con le rappresentazioni sociali e sollevando questioni di natura teorica sui rapporti tra soggettività, dispositivi di misura e governance. Studiando il passaggio dall'ascolto fenomenologico a quello che verrà definito *ascolto operazionale* come pratica di misurazione e interpolazione statistica, l'analisi tenterà di decostruire la narrazione della "città senziente" al cuore del concetto stesso di *sensibilità* che quei dispositivi mettono in campo, determinando una precisa definizione di intelligenza (artificiale) con cui la retorica "smart" sembra collidere.

L'impostazione metodologica di questo studio pensa, costruttivisticamente, gli oggetti

tecnologici non come meri strumenti impiegati per fini operativi teleologicamente determinati (l'ordine, l'efficienza, la qualità), ma come costruzioni attraverso cui la società definisce sé stessa. Le macchine e gli artefatti tecnici incorporano, nei loro codici e principi di funzionamento, i saperi e i valori che delimitano l'orizzonte epistemico della società contemporanea, di cui la smart city, in quanto piattaforma sorretta da quei sistemi, è rappresentante privilegiato. Gli algoritmi sono, in questo senso, oggetti sociali a pieno titolo e vanno studiati in maniera "archeologica". In un rapporto circolare tra tecnologie, saperi e società, gli algoritmi incorporano epistemologie e pratiche culturali determinate dall'interazione tra le loro stesse possibilità tecniche e gli attori umani che li programmano, mentre a livello socio-economico ed ideologico l'adozione di quegli stessi sistemi è giustificata discorsivamente e incentivata monetariamente. Se l'introduzione di dispositivi di ascolto macchinico in ambiti sempre più estesi è promossa a livello retorico, facendo leva sulla fascinazione della macchina che comprende il parlato o che riconosce i suoni – un vero e proprio tropo della cultura occidentale, come ricostruito da Trevox Cox (2018) –, essa fa capo anche a un'altra logica, fondata sulle potenzialità del *data mining* di estrarre valore da una raccolta dati quanto più ramificata possibile (Srnicek, 2017). Mentre il *riconoscimento vocale* seduce gli utenti con la possibilità di parlare con i dispositivi in maniera naturale (Norman, 2010; Dolar, 2014), come in un faccia-a-faccia che umanizza la macchina e crea fiducia, esso da quelle interazioni estrae dati e profili biometrici che sono potenzialmente implementabili in ambiti sensibili quali l'identificazione del parlante e l'individuazione di parole chiave (*keyword spotting*) nelle intercettazioni ambientali, con applicazioni in ambito sia securitario che forense. Lo stesso principio è valido per i nuovi sistemi di *audio monitoring* nelle smart city, in grado di identificare traffico, rumori sospetti, situazioni di emergenza e pericolo, ponendosi al contempo come dispositivi di sicurezza e canali di intrusività. In questo scenario l'ascolto macchinico gioca un ruolo strategico, in quanto, in virtù del suo potere seduttivo, da un lato opera una rappresentazione "umanizzante" e dunque più accettabile dell'intelligenza artificiale, mentre dall'altro permette di insinuare i sistemi algoritmici in ambiti sempre più estesi, nelle case, nelle città, negli uffici, nell'ambiente in generale, allargando il campo della raccolta dati e fornendo dunque più

materiale per un addestramento macchinico (*machine learning*) sempre più preciso. Questa modalità di ascolto ubiquo e permanente, di importanza strategica per quell'*ubiquitous computing* (Greenfield, 2006) che è alla base del modello "smart", si profila come paradigma di un nuovo tipo di *sorveglianza* e invita a una riflessione sulla questione delle *privacy*.

Coerentemente con questa impostazione, partirò dall'analisi dei sistemi di ascolto macchinico, ovvero dal "come ascoltano le macchine" e dal come, questa particolare modalità di "ascolto", che consiste in funzioni di misurazione, classificazione e confronto, contribuisca a definire tanto la smart city quanto i suoi soggetti, inaugurando nuove pratiche socio-comunicative.

2. L'ascolto nella smart city

Nelle rappresentazioni recenti la smart city è descritta sempre più come «città senziente», un organismo in grado di sentire e regolarsi attraverso organi artificiali. Sensori, videocamere, sistemi di rilevamento costituiscono gli occhi e le orecchie di questo grande apparato. Questo tipo di retorica ricorre sia nei materiali di marketing (Delale, 2019), che nelle pubblicazioni scientifiche, sia di natura tecnica (Doran, Gokhale e Dagnino, 2013) che critica (Chopplet, 2018; Garnier, 2019). Un perfetto esempio di retorica "smart" basata sulla metafora organicistica è riscontrabile nel progetto Smart Dubai, che prevede una massiccia implementazione di IoT nella città di Dubai destinati al monitoraggio delle più disparate funzioni della città: «Every city has a heart and soul of its own and may claim same rights as a human being. If we are to recognize the entire city to be a living entity, just like a smart human being, a smart city would also require constant monitoring of its health» (Sahib, 2020, p. 437). Tra le modalità di monitoraggio costante della «salute» di questo organismo, un ruolo di rilievo viene riservato all'ascolto, sotto forma di monitoraggio sonoro del traffico, dell'inquinamento acustico e dei suoni di arma da fuoco. Sensori acustici sono impiegati per rilevare colpi di pistola e attivare automaticamente risposte d'emergenza quali

il lampeggiamento delle luci stradali circostanti. Lo stesso sistema è impiegato per monitorare l'inquinamento acustico urbano sia in relazione al traffico che alla vita notturna.

Se la metafora organicistica ricorre sin da tempi remoti (Mumford, 2002), con la smart city tutto ciò assume un valore letterale e, per così dire, incarnato. La città intelligente "ascolta" per davvero, capta segnali sonori, li registra, li misura, li confronta. Utilizzando una tecnologia che deriva dal riconoscimento vocale e dal riconoscimento automatico del parlatore (Roe e Wilpon, 1994), quale quella impiegata dai dispositivi domotici in voga negli ultimi anni (Amazon Alexa, Google Home, etc), l'*audio monitoring* può ampliare il campo di captazione dall'ambito domestico all'intero ambiente sociale.

EAR-IT (*Experimenting Acoustics in Real environment using Innovative Test-beds*) è un progetto sviluppato nel 2012 da EGlobalMarket, azienda francese che fornisce servizi di analisi dati, in collaborazione con numerosi istituti di ricerca sulle nuove tecnologie[1], nell'ambito del progetto europeo FIRE SmartSantander. Il progetto vede la città di Santander e quelle del suo network (Belgrado, Guildford, Lubeck) fare da banco di prova per nuove tecnologie da applicare alle smart cities: le città diventano piattaforme per la sperimentazione delle IoT (*Internet of Things*) e in particolare per l'applicazione di reti di sensori audio sia in ambienti chiusi che all'aperto, finalizzate alla creazione di una «intelligenza distribuita alimentata acusticamente» (CORDIS, 2017).

Il design del sistema EAR-IT prevede l'impiego di sensori acustici APU (*Acoustic Processor Units*) costituiti da microfoni, processori atti a ottimizzare il segnale, e algoritmi in grado di riconoscere in quel segnale gli elementi che lo connotano come un determinato evento acustico sensibile, oppure come un determinato comando vocale (Pham e Cousin, 2013). Gli APU «ascoltano continuamente» l'ambiente e pre-analizzano il suono localmente e in tempo reale; essendo distribuiti in maniera ramificata essi sono in grado di comunicare tra loro per localizzare con precisione l'evento sonoro; in caso di riscontro mettono in atto automaticamente l'intervento per cui sono stati programmati.

Il sistema EAR-IT è stato testato prima di tutto sul traffico di Santander: in un incrocio

[1] Università di Cantabria (Spagna), Lulea (Svezia), Uninova (Portogallo), l'organizzazione degli istituti di ricerca applicata Fraunhofer (Germania), Wuxi Smart Sensing Company (Cina) e la Fondazione Mandat (Svizzera).

particolarmente trafficato in cui i veicoli di emergenza hanno difficoltà a farsi strada, il sistema di sensori acustici dislocati lungo le strade cittadine e comunicanti tra loro, riconosce il suono delle sirene determinandone anche provenienza e direzione e interviene cambiando automaticamente i semafori in favore del veicolo di emergenza (Commissione Europea, 2014). Lo stesso sistema viene impiegato per monitorare il traffico in base ai livelli di rumore, o per rilevare situazioni di pericolo, come un grido di aiuto o uno sparo; in questi casi il sistema può allertare automaticamente le autorità senza passare per il vaglio dell'operatore umano (Euronews, 2014).

Le tecnologie di *acoustic monitoring* sono impiegate, parallelamente all'ambito cittadino, anche in quello domestico e sanitario. L'Istituto Fraunhofer di Oldenburg, che ha contribuito al progetto EAR-IT, ha brevettato anche il prototipo SonicSentinel pensato per gli anziani, in grado di riconoscere suoni legati a situazioni di emergenza come cadute, tosse forte, urla o lamenti, inviando segnali ai sanitari e alla famiglia per sollecitarne l'intervento (Fraunhofer IDMT, 2014)[2]. Il servizio Alexa Guard, nuova implementazione di Amazon Echo, rileva suoni sensibili per la sicurezza della casa quali vetri rotti e allarmi antincendio e fa scegliere agli utenti quale tipo di operazione mettere in atto al rilevamento: invio di notifica sul telefono, collegamento con il microfono di Echo per ascoltare i suoni della casa in tempo reale, lampeggiamento delle luci domestiche o segnale di allarme (https://www.amazon.com/b?ie=UTF8&node=18021383011, 04/09/2020).

Un altro caso studio di particolare interesse è il progetto SONYC (Sounds of New York City), sviluppato dalla New York University (https://wp.nyu.edu/sonyc/, 04/09/2020; Bello *et al.*, 2019). Il progetto prevede l'implementazione, nella città di New York, di un sistema ramificato di sensori audio deputati sia alla misurazione che al riconoscimento del rumore cittadino, con l'obiettivo di individuare *pattern* significativi di inquinamento acustico, in modo da poter adottare misurare mirate e ottimizzate per ridurlo. «SONYC utilizes big data

[2] Nello stesso centro è stato messo a punto un sensore acustico che monitora le condizioni di una macchina industriale attraverso i suoni che produce, permettendo di individuare malfunzionamenti nel momento stesso in cui si manifestano, riducendo, dunque, i tempi di intervento e ripristino (Fraunhofer IDMT, 2019; Nirjon, Srinivasan e Sookoor, 2017). Inoltre, il nuovo sviluppo di EAR-IT da parte della Fraunhofer riguarda anche la sicurezza e l'efficienza energetica degli edifici. Il sistema è in grado di rilevare, attraverso il monitoraggio audio, la presenza di individui nelle stanze, identificando «chiusura di porte, ticchettii sulle tastiere, macchine del caffè e distributori in azione», in modo da poter attivare e disattivare la corrente nelle stanze inutilizzate (Kelly *et al*., 2014, p. 661).

solutions to analyze, retrieve and visualize information from sensors and citizens, creating a comprehensive acoustic model of the city that can be used to identify significant patterns of noise pollution» (Bello *et al.*, 2019, p. 68). La particolarità del progetto SONYC risiede nella creazione di un network socio-tecnico che include sia sensori che cittadini, adottando una modalità di "collaborazione" tra umani e macchine paradigmatica per l'interpretazione critica che si intende qui argomentare. Laddove, infatti, i 55 sensori distribuiti sul territorio cittadino inviano segnali audio al server, dove vengono analizzati in tempo reale, la "comprensione macchinica" di quei segnali dipende in larga misura dal contributo dei cittadini, i quali possono prendere parte al progetto attraverso una piattaforma online per "etichettare" frammenti delle registrazioni sonore (https://www.zooniverse.org/projects/anaelisa24/sounds-of-new-york-city-sonyc, 04/09/2020). Attraverso il loro contributo, i cittadini aiutano ad alimentare gli algoritmi di *machine learning supervisionato*[3] su cui si basano i processi di analisi di SONYC: segnalando la corrispondenza tra un segnale audio e una classe di suoni, l'intervento "insegna" qualcosa di nuovo all'algoritmo, permettendogli di migliorare la sua capacità di riconoscimento. Come verrà spiegato nei prossimi paragrafi, è proprio la categoria di "riconoscimento" a ricoprire un ruolo chiave nel *machine learning*. Qui la collaborazione tra umani e algoritmi, sebbene rappresentata spesso a livello discorsivo e immaginario come un processo di umanizzazione della macchina, si realizza piuttosto in uno spostamento della posizione dell'umano da modello per la macchina a *medium* per il funzionamento di questa, come accade per l'intervento dei cittadini nel caso appena preso in esame.

Dai casi studio passati in rassegna, e dalle tecnologie in essi impiegate, emergono tre punti significativi su cui vorrei concentrare l'analisi: *ascolto operazionale*, *intervento*

[3] Nell'ambito del *machine learning* si distingue tra tecniche "supervisionate" e "non supervisionate". Con il primo termine si fa riferimento ad algoritmi cui vengono forniti in input, in fase di *training*, dati "etichettati", ovvero classificati; qui il compito dell'algoritmo è "apprendere" la relazione tra dati e classi e generalizzarla a nuovi casi (ad esempio, data una serie di immagini e data una mappatura tra alcune di esse e la classe "gatti", l'algoritmo sarà in grado di individuare i gatti anche in immagini non presenti nel suo training set). Le tecniche di *learning* "non supervisionato", invece, prevedono in input dati non etichettati; il compito dell'algoritmo è qui quello di "classificare" i dati senza avere alcun tipo di riferimento, ma affidandosi unicamente alle differenze, le ricorrenze e le correlazioni che esso è in grado di rintracciare tra i dati. Per un approfondimento sul tema si veda il completo volume di Goodfellow, Bengio e Courville (2016), esaustivo sulla logica generale dell'apprendimento macchinico, dal *machine learning* alla sua più recente evoluzione *deep* a base di reti neurali.

emergenziale e *invisibilità*; tre principi di funzionamento dell'ascolto macchinico che contribuiscono a ridefinire le modalità della comunicazione, del controllo sociale e della privacy nella smart city.

3. Ascolto operazionale e intervento automatico

Tutti i sistemi di audio monitoring sopra menzionati puntano a produrre una *risposta automatica* all'evento sonoro riconosciuto, una immediata traduzione in atto: al riconoscimento di un'ambulanza cambia l'orientamento dei semafori, a quello di uno sparo lampeggiano le luci e si allertano le autorità, a quello di vetri rotti si invia una notifica e così via. Qui risiede il fulcro di ciò che propongo di definire *ascolto operazionale*, in cui l'output di alcune operazioni è un'altra operazione, un'operazione automatica che bypassa il momento intellettuale della comprensione, dell'«efficienza simbolica» (Zizek, 2012), per intervenire direttamente sul reale. Questa modalità invita a riflettere sul passaggio epistemologico prodotto dalle tecnologie algoritmiche, che insieme all'ascolto modificano il concetto stesso di *conoscenza*.

L'utilizzo del termine "ascolto operazionale" deriva da un suggerimento di Mark Andrejevic (2019, p. 108), il quale utilizza il concetto di «operational images» in contrapposizione a quello di rappresentazione: laddove quest'ultimo riguarda immagini rivolte allo sguardo ermeneutico, le immagini operazionali sono traduzioni in forma visibile di operazioni macchiniche non destinate allo sguardo umano. In maniera analoga, invito a pensare l'ascolto operazionale come una modalità prettamente macchinica, che tuttavia si esprime con effetti tangibili e interpretabili dall'umano sulla base di analogie con il proprio apparato uditivo. L'utilizzo del termine "operazionale" ha tuttavia anche un fondamento sociologico nel lavoro di Marcuse, il quale parlava dell'«operazionalismo» come una forma di castrazione del potere speculativo del pensiero e sua riduzione alla dimensione strumentale come unico orizzonte di senso. Ciò si manifesta in primo luogo nell'impoverimento del linguaggio, laddove «il significato […] viene ristretto alla

rappresentazione di particolari operazioni» (Marcuse, 1999, p. 26), ma produce conseguenze ontologiche oltre che semantiche, poiché «tende ad identificare le cose con la loro funzione» (p. 98). La monodimensionalità della ragione strumentale si traduce, per Marcuse, nell'uso di concetti operativi che non necessitano comprensione o credenza, ma solo conferma nell'azione, azione conforme. Questa impostazione critica può essere d'aiuto anche per comprendere le tecnologie algoritmiche. Il computer, infatti, non "è" ma "fa", non rappresenta, ma produce effetti, non descrive ma «inscrive/scrive dentro il mondo» (Floridi, 2017, p. 162).

L'operazionalismo dei media automatici, con e oltre Marcuse, ha a che fare con la coincidenza di comprensione e operazione, che trasforma la conoscenza in una pratica di *riconoscimento*. Questo è un aspetto decisivo per comprendere anche la peculiarità dell'ascolto macchinico. La macchina non ascolta in maniera fenomenologica, ovvero il suo ascolto non è modellato su quello umano, non è, come spiegherò a breve, inscindibilmente legato all'intenzionalità, all'esperienza e alla memoria. Ciò che la macchina fa quando ascolta è, tecnicamente, una *misurazione* di valori e una *classificazione* di questi secondo parametri da essa stessa stabiliti, utilizzando tecniche di "apprendimento automatico" (*machine learning*). Con questo si intende che gli algoritmi di *machine learning* non operano più sulla base di modelli pre-determinati dei fenomeni, ma sono in grado di trovare, immanentemente ai dati stessi, le rappresentazioni più idonee di quei fenomeni in base alle necessità e propensioni macchiniche (Goodfellow, Bengio e Courville, 2016)[4]. Gli algoritmi vengono "addestrati" a riconoscere eventi sonori di un certo tipo attraverso l'analisi di una grande quantità di dati audio di eventi sonori simili (il cosiddetto *training dataset*) e "apprendono" le caratteristiche fondamentali dello spettro sonoro degli eventi in esame attraverso la classificazione statistica di parametri quali distribuzione di frequenza, inviluppo, picchi – i cosiddetti *landmarks* e *fingerprints* (Wang, 2003; Pieraccini, 2012;

[4] Questo approccio è stato messo a frutto in particolare dal *deep learning*, particolare tipo di *machine learning* che, al posto di sistemi statistici quali Hidden Markov Model, impiega reti neurali artificiali per effettuare classificazioni "in profondità", ovvero su molteplici livelli gerarchicamente distribuiti in base al dettaglio della classificazione. Sebbene la logica soggiacente sia quella del *machine learning*, Goodfellow e altri mettono in evidenza come l'approccio *deep* sia particolarmente efficace in ambiti non deterministici e variabili, quali appunto l'ascolto e la visione, in cui i dati sono difficilmente modellizzabili (Goodfellow, Bengio e Courville, 2016).

Hollosi *et al*., 2013).

L'apprendimento algoritmico consiste, dunque, in un riconoscimento di *pattern* appresi durante processi di "training" su base dati, e nella capacità di generalizzare, ovvero di applicare i parametri di classificazione utilizzati negli esempi di addestramento a nuovi casi, non visti prima. In questo modo la macchina è in grado di riconoscere una classe di suoni, con molte delle sue possibili varianti – dettaglio decisivo, vista l'estrema variabilità del campo sonoro, che ha decretato il successo del *machine learning* su altri sistemi[5].

L'adozione di sistemi di apprendimento automatico a base di dati (*data-driven*) ha determinato un cambio di paradigma nell'ambito dell'intelligenza artificiale, spostando gli interessi di ricerca da un'intelligenza «produttiva», in grado di simulare i processi cognitivi umani, a un'intelligenza «riproduttiva», che non ha, ormai, molto in comune con l'intelligenza umana, ma ne «emula gli effetti» (Floridi, 2017, p. 159), basandosi unicamente sull'analisi di grandi quantitativi di dati presi dall'esistente. In questo quadro è l'atto del comprendere che cambia significato, assimilandosi sempre più a un riconoscimento, il quale si esprime attraverso un'azione diretta sul reale che annulla la distanza simbolico-linguistica.

L'ascolto operazionale, dunque, si può definire come un processo macchinico che parte dal rilevamento del suono e, a seconda del tipo di riconoscimento a seguito di calcoli specifici, si tramuta immediatamente in azione. Questa azione può essere l'attivazione di un segnale, la chiamata, il report, ma anche la semplice memorizzazione o cancellazione dell'audio.

3.1 Ascolto fenomenologico e ascolto culturale

Le teorie dell'ascolto dell'ultimo secolo hanno adottato spesso un approccio fenomenologico per descrivere la specifica modalità di conoscenza messa in atto dal senso

[5] Questa tecnologia è stata implementata prima nell'ambito del riconoscimento vocale, e successivamente negli algoritmi di *copyright detection* come ContentID usato da Youtube e Facebook, nonché in Shazam e servizi simili di *audioprint*.

dell'udito. Queste descrizioni mettono in evidenza la natura specificamente intima e immersiva dell'ascolto, che ne fanno una facoltà essenzialmente esperienziale, in contrasto con la vista come organo di misura oggettiva e distanza.

«The listener is entwined with the heard. His sense of the world and of himself is constituted in this bond» (Voegelin, 2010, p. 5). Per l'approccio fenomenologico, l'ascolto non *descrive* ma *produce* il fenomeno acustico; l'oggetto sonoro *non è* senza l'ascolto, è tale solo in quanto ascoltato. L'ascolto fenomenologico è immediatamente tradotto in esperienza, produce un rapporto com-prensivo e senza distanza tra il soggetto e il mondo, in cui sia il soggetto che l'oggetto divengono tali nell'atto dell'ascoltare da parte della coscienza intenzionale. Per l'ascolto fenomenologico il suono è il "correlato" dell'intenzionalità della funzione psichica (Barthes, 2001).

Rifacendosi a un'osservazione di William James secondo cui «when we listen to a person speaking – much of what we think we hear – is supplied from our memory», François Bonnet sottolinea come il suono lasci sempre una traccia nella memoria: «to perceive is always to immediately remember having perceived», percepire è sempre immediatamente ricordare di avere percepito (Bonnet, 2016, p. 75). Senza ritenzione mnestica il fenomeno sonoro svanirebbe nel momento stesso del suo apparire, invalidando tanto l'idea di ascolto quanto quella di comunicazione orale: nulla rimarrebbe dei suoni delle parole oltre il momento esatto della loro vibrazione. È in questo rinvio, in questa latenza, che l'ascolto fenomenologico si apre a un esterno fatto di segni e tracce che già profilano un mondo macchinico, artificiale, quello della ritenzione terziaria, dei supporti, delle tecniche.

Qui risiede il momento di inscindibile fusione tra un ascolto fenomenologico e un ascolto "culturale", ovvero un ascolto socialmente determinato dall'insieme di saperi e tecniche inscritte negli artefatti che ne determinano la permanenza. L'ascolto avviene sempre tra il soggetto, la sua storia, le sue aspettative e i dispositivi socio-tecnici. La definizione fenomenologica di un soggetto che, ascoltando, esperisce il mondo «senza distanza» (Voegelin, 2010, p. 5), è anche una nozione astratta di ascolto, quella di un "ascoltatore zero" senza storia e decontestualizzato; essa si trova così di fatto decostruita dalle esteriorizzazioni necessarie al suo stesso prodursi. Il suono non sarebbe un "fenomeno" per

il soggetto senza lasciare tracce nella sua coscienza, ma, affinché vi siano tracce, la relazione tra soggetto e mondo deve già essere esteriorizzata, mediata da dispositivi di iscrizione e supporto di quelle tracce. È per questo che l'ascolto fenomenologico non è mai puro, ma sempre integrato con altri tipi di ascolto, incluso l'ascolto macchinico. Dire che le macchine "ascoltano" è una metafora, in quanto esse non ascoltano come ascoltiamo noi; ma è il nostro stesso ascolto a non coincidere pienamente con la sua definizione, a essere continuamente influenzato e riconfigurato dal suo fuori, dagli strumenti tecnici e dai saperi in essi incorporati. Parlare di ascolto, dunque, implica sempre la necessità di rivolgersi ai due lati della medaglia, all'umano e ai suoi strumenti. La fenomenologia presuppone sempre l'epistemologia, la cultura, i rapporti di potere (Sterne, 2003, p. 13).

3.2 Ascolto macchinico, media, soggettività

L'ascolto operazionale si discosta dall'ascolto fenomenologico in diversi punti: in primo luogo, misurando il suono con parametri oggettivi, esso postula l'esistenza del mondo senza il soggetto che lo esperisce e lo costituisce, in quanto coglie il suono al di fuori del suo darsi alla coscienza; in secondo luogo è *desoggettivato*, poiché la macchina che ascolta non opera alcuna fusione col suo oggetto, piuttosto entifica il suono, non lo "com-prende", ma lo tiene a distanza, se ne appropria secondo parametri macchinici che nulla più hanno a che fare con le modalità umane, rompendo il legame ascolto-comprensione-soggettività.

Se molti dei timori legati all'ascolto macchinico derivano da una fuorviante rappresentazione di questo sul modello dell'ascolto umano, è altrettanto vero che non esiste un "ascolto umano" in quanto tale, in quanto esso è sempre completato da organi artificiali, media, protesi, che lo ridefiniscono e riconfigurano (Sterne, 2003). Per dirla col Marx dei *Manoscritti*, «l'educazione dei cinque sensi è un'operazione di tutta la storia del mondo», i sensi umani sono coltivati e formati socialmente, tanto dai saperi quanto dagli artefatti (Marx, 2004, p. 114). Per Mauss il corpo è il primo oggetto tecnico dell'uomo (Mauss, 2017), e così anche l'ascolto, investito, in particolare dopo l'avvento della registrazione

sonora, da nuove possibilità socio-tecniche di esercizio e definizione, è storicamente determinato. La modellizzazione dell'ascolto, nella sua modalità organica, è presa in un circolo biotecnico in cui organi e protesi (fonografi, registratori, microfoni…) sono intrecciati, e contribuiscono alla modificazione del percepito, del sistema uditivo, e dei modi d'ascolto stessi. Come sottolinea Peters, i media sonori non solo imitano l'organo uditivo, ma nel farlo tentano anche di plasmarlo come un determinato strumento, «metaforizzano» quello stesso organo mostrandone la natura artefattuale (Peters, 2004, p. 183). Il fonografo, ad esempio, come cristallizzazione di un insieme di nuovi saperi medici e fisici circa l'apparato uditivo e la risonanza, si pone come "protesi" acustica modellata sul funzionamento dell'orecchio (Sterne, 2003); ma nel contempo ridefinisce anche l'ascolto preparandolo a un nuovo tipo di comunicazione, quello della telefonia, in cui il faccia-a-faccia cede il passo alla presenza dislocata e differita (Borrelli, 2000; Ernst, 2016).

La misurazione "oggettiva" del suono tramite strumenti (spettrografi e analizzatori) è un altro passo verso la desoggettivazione, poiché scinde il percepito dal fenomeno vibratorio in sé, inaugurando quindi un nuovo modo di ascolto che è a metà tra il fenomenologico e il macchinico, un ascolto che media tra ciò che i sensi percepiscono e ciò che le macchine impietosamente misurano. È l'ascolto dei tecnici del suono, dei compositori di musica acusmatica, che trattano il suono "in quanto tale", unicamente per le sue proprietà morfologico-vibratorie, separandolo dalla sua causa naturale (un approccio che è stato definito «ascolto ridotto», Schaeffer, 2017).

I media computazionali e di rete modificano ulteriormente l'idea di ascolto e comunicazione, poiché alla presenza differita sostituiscono la totale disincarnazione dei processi algoritmici. L'ascolto che accade sul *cloud* è infatti un ascolto che, avvenendo in un luogo esterno e insondabile rispetto tanto all'individuo quanto al corpo sociale, prescinde da qualunque corporeità, non ha necessità né di soggetto parlante né di soggetto ascoltante. Il suono come strumento ermeneutico per misurare la vita di una città, oppure la salute di macchine industriali ed edifici, un suono prodotto da macchine e ascoltato da altre macchine, prescinde completamente dalla presenza umana e dalle sue modalità percettive. L'umano passa, in questi due momenti epistemici, da "soggetto ascoltante" a "modello" per

l'ascolto macchinico, e infine perde anche quella posizione di modello per divenire *medium* nella comunicazione tra macchine, fornitore di dati per alimentare algoritmi che non condividono più nulla con i suoi apparati uditivi. Luciano Floridi ha definito questo fenomeno «computazione basata sull'uomo» (Floridi, 2017, p. 168).

Questo passaggio pare confermato dal caso SONYC, nonché da molteplici altri contesti in cui gli individui scelgono volontariamente di fare da tramite per il lavoro macchinico, come in Amazon Mechanical Turk. La chiave della tecnologia SONYC non è tanto che il suono venga misurato, cosa che avveniva già con i dispositivi analogici di rilevamento dell'impatto acustico, quanto che esso venga "riconosciuto" e discriminato, a seconda che si tratti di traffico, sirene, cantiere, spari e così via. È qui che le possibilità offerte dal *machine learning* segnano una differenza tanto sociale quanto epistemologica: l'intervento umano diventa gregario di un processo macchinico con le sue specificità e le sue esigenze, in cui il riconoscimento è finalizzato all'operazione automatica.

L'ascolto operazionale del *machine learning* non "imita" l'umano, bensì scioglie ogni legame col modello dell'apparato uditivo, col modello "timpanico" del fonografo, per mettere in atto il modo specificamente macchinico di trattare i dati acustici (Li, 2017). È la macchina a scegliere gli elementi significanti del segnale in base alle sue classificazioni numeriche interne[6]. Il momento decisivo nella storia dell'IA, che ha determinato il successo dell'approccio *data-driven* su quello *model-based*, è infatti coinciso con il riscontro che la macchina potesse trovare da sé le proprie rappresentazioni, i propri parametri utili, all'interno dei dati; parametri che non coincidono né con le formalizzazioni linguistiche né con le strutture simbolico-culturali.

Ciò ha due conseguenze importanti: in primo luogo il problema dell'ascolto viene a riconfigurarsi come un problema generale di *pattern recognition*, di trattamento matematico dell'informazione applicabile a qualunque campo di dati, dal sonoro, al fisico, al sociale; in

[6] Questo paradigma, che si è determinato negli anni '80 nei laboratori IBM per risolvere il difficile problema del riconoscimento automatico del parlato, si basa sul semplice assunto che le macchine fanno le cose in maniera diversa rispetto agli umani (ad esempio gli aeroplani non battono le ali per volare). Analogamente, non è necessario comprendere "come l'uomo parla" per approdare a risultati di riconoscimento vocale, ma è piuttosto opportuno chiedersi "come la macchina ascolta", spostare dunque il problema dal trasferimento computazionale della teoria linguistica, all'analisi di dati di parlato "reale" senza teoria (Li, 2017).

secondo luogo l'ascolto macchinico perde ogni analogia con l'ascolto umano, passando dal fisiologicamente percepibile al numericamente ottimale, dall'udibile al misurabile, sciogliendo dunque ogni legame tanto col senso quanto con la "comprensione"[7]. La macchina può *riconoscere* le parole o i suoni senza *comprenderne* il significato, passando direttamente dal riconoscimento all'azione. Analogamente può riconoscere contenuti da censurare senza veramente comprenderne il senso, come negli algoritmi di Facebook, o le probabilità di azioni criminali, senza comprenderne le cause socio-culturali, come negli algoritmi di *predictive policing* (Mohler, 2015). Il rapporto tra ascolto, comunicazione e comprensione è trasformato da questa modalità operativa in una direzione prettamente matematica, e, sull'impronta della teoria di Shannon, trascura il messaggio per rivolgersi unicamente al trattamento del segnale (sonoro o di altro tipo) e al suo trasferimento da macchina a macchina (Mills, 2012).

Ascolto fenomenologico e ascolto operazionale non vanno, dunque, pensati come sinonimi di ascolto umano e ascolto macchinico, ma come due modalità che si intrecciano ed integrano continuamente tra l'umano e il macchinico. L'ascolto è qualcosa che accade *tra* umani e macchine ed è continuamente ridefinito socio-tecnicamente.

4. Emergenza, naturalizzazione, sicurezza

Tutti i sistemi di ascolto automatico presi qui in esame riservano grande attenzione al *fattore tempo*, ovvero alla possibilità di ridurre, a seconda dei casi, il tempo di percorrenza (come nel caso delle ambulanze negli incroci trafficati), il tempo di intervento (come nel caso degli spari o delle cadute di anziani), o il tempo di attività/inattività (come nel caso del *downtime* delle macchine industriali o dello spreco di energia negli uffici vuoti). Ciò pare confermare un paradigma *emergenziale* delle tecnologie smart, in cui l'intervento puntuale

[7] È pur vero che, affianco all'ascolto automatico, un grosso lavoro di ricerca è condotto sul *natural language understanding* (NLU), che correla statisticamente il parlato automaticamente riconosciuto con possibili ordini semantici. Tuttavia questo processo – che per alcuni fa accedere la macchina a un livello di proto-comprensione (Floridi, 2017, p. 147) – non avviene immanentemente all'ascolto, ma solo successivamente, a livello della traduzione del suono in testo.

just-in-time a un evento specifico sostituisce l'intervento di lungo periodo con strumenti amministrativi e politici. In un quadro più ampio, è l'amministrazione stessa ad abdicare in favore degli interventi automatici messi in atto da macchine algoritmiche. Come messo in risalto da Morozov (2017), ciò pare essere un paradigma distintivo della smart city che rivela un tratto di depoliticizzazione, incentivato, se non addirittura estorto, dalle grandi aziende tecnologiche che ambiscono a sussumere quanti più aspetti della vita pubblica nei loro servizi privati. Ciò va di pari passo con una *naturalizzazione* della sfera sociale che deriva dall'epistemologia introdotta dai sistemi algoritmici basati su dati, che, come visto, trattano ogni segnale a livello matematico-statistico, al pari di un evento fisico o naturale di cui è possibile prevedere il comportamento correlando i dati delle condizioni iniziali[8].

Il sapere operazionale volto all'azione automatica, inaugurato dall'ingresso degli algoritmi nei più vari campi del sociale e del politico, prende sempre più il posto del sapere intellettuale votato alla comprensione dialettica e alla ragione argomentativa tipiche dell'agire comunicativo di stampo umanistico (Habermas, 1981). Esso mette dunque in collegamento diretto le modalità statistiche della governance con una nuova configurazione securitaria della società, in cui la predizione dell'evento critico è più importante della comprensione delle sue cause, e in cui l'intervento puntuale ed emergenziale prende il posto dell'intervento politico di lungo periodo. Ciò che definisce un paradigma securitario, infatti, non è tanto la correzione delle devianze quanto la previsione delle condotte, la conoscenza anticipata di ciò che può accadere per poterlo evitare prima che accada. I modelli di simulazione algoritmica a base di dati non solo permettono tale tipo di conoscenza, ma lo fondano epistemologicamente.

Traducendo ogni fenomeno sociale in dati misurabili, i sistemi *data-driven* riconducono la differenza qualitativa a "misura comune" nell'ambito di griglie e classifiche di valutazione della performance. Il *ranking* diviene uno strumento di "naturalizzazione" che, enfatizzando il merito individuale, denega l'origine sociale della disuguaglianza (Borrelli, 2015). La *valutazione* si rivela così essere un dispositivo governamentale della società

[8] Non è un caso che gli algoritmi di «predictive policing» della polizia statunitense siano ispirati a quelli per il monitoraggio dei terremoti (Mohler, 2015).

digitale, quella che Rouvroy definisce «governamentalità algoritmica» (Rouvroy e Berns, 2013), una modalità di governo senza norme e immanente al campo sociale. Come nota Tarizzo, risiede qui il passaggio da una società disciplinare a una «società etopolitica», una società dell'ottimizzazione in cui i comportamenti umani «sono svincolati da ogni centro soggettivo di imputabilità […] sono cioè segmentati e misurati nella loro specifica operatività, per essere poi adeguatamente potenziati, diversificati e flessibilizzati» (Tarizzo, 2013, p. 51). La società etopolitica non punisce i *soggetti*, ma valuta quell'assemblaggio composito di dati che è il loro profilo per poter prevenire gli *eventi*; e al fine di una valutazione sempre più accurata si nutre di ogni possibile dettaglio, inclusi casi particolari, diversità, devianze, per far funzionare la sua macchina includente.

5. Invisibilità e privacy

Coerentemente con questa analisi, è possibile notare come tutte le tecnologie di cui sopra puntino sull'invisibilità e la discrezione del campo sonoro per legittimarsi nei confronti dei loro destinatari e, di conseguenza, estendere il campo della raccolta dati in ambiti sempre più sensibili. A differenza della videosorveglianza, l'*acoustic monitoring* viene meglio accolto dagli utenti e dall'opinione pubblica (Hollosi *et al*., 2013, p. 339). Juliette Volcler nota come la crisi di legittimità prodotta dalla diffusione mediatica della violenza della guerra del Vietnam fu risolta proprio dall'introduzione di armi soniche, in grado di colpire invisibilmente e senza spargimento di sangue, per quanto non meno cruentemente (Volcler, 2012). È proprio la discrezione a rendere l'audio monitoring un perfetto strumento di sorveglianza ubiqua, che usa quella apparente non-intrusività per insinuare la tecnologia smart in maniera sempre più capillare, dai luoghi pubblici a quelli di lavoro a quelli privati e domestici.

Ad esempio, affinché il sistema SONYC sia in grado di individuare *pattern* significativi, ha bisogno di incrociare i dati sonori con altri dati, quali la mappa dei locali notturni, dei punti di aggregazione sociale o turistica, nonché dati provenienti dai social media e della

geolocalizzazione personale (Bello *et al*., 2019), in modo da comprendere con precisione su quali fattori concentrare l'intervento automatico. Ma, ricostruendo le correlazioni tra movimenti delle persone, punti di incontro e ambiente sonoro, il sistema viene in possesso di una quantità di dati sensibili tale da descrivere in maniera molto dettagliata i comportamenti di una parte della società. Facendo leva sul tema del miglioramento dell'inquinamento acustico e sulla discrezione dei suoi sensori, il sistema diventa parte integrante di un'infrastruttura di estrazione dati trasversale e ramificata, che non si limita al solo ambito acustico, ma interconnette informazioni di ogni sorta.

In generale, pare qui evidente il passaggio da una logica del controllo individuale a una logica del monitoraggio «ambientale» (Andrejevic, 2019, p. 85): l'enfasi non è più sul governare i soggetti ma sul prevedere eventi, a cui tutta la popolazione è interessata come campo statistico. La sorveglianza si sposta dal soggetto all'intero ambiente sociale, attraverso l'impiego di sensori che registrano dati in maniera orizzontale e immanentemente, senza concrezioni di valore o di senso a priori. I dati non sono estratti con un fine, e per questo tutti i dati servono; saranno gli algoritmi, poi, a far emergere il senso dalla correlazione dei dati stessi, un senso che, per definizione, sfugge alla sensibilità umana poiché frutto dell'elaborazione di una quantità troppo elevata di informazioni e variabili. Qui la diversità è una risorsa generativa: quanto più vari i dati, tanto più accurati i modelli di previsione e valutazione. Ed è l'intera popolazione, l'ambiente stesso, la fonte di questa varietà.

In questo quadro, la sorveglianza diviene un paradigma epistemico, quasi una condizione tecnica, prima che un dispositivo di controllo sociale (Zuboff, 2019). Ciò che Andrejevic chiama «post-panopticon» (Andrejevic, 2019, p. 75) è infatti il passaggio dallo spettacolo deterrente della sorveglianza come processo di soggettivazione, alla scomparsa della sorveglianza e della sua efficienza simbolica in funzione della sua ubiquità. «If the disciplinary goal is the spectacle of surveillance, the pre-emptive one is its disappearence through ubiquity. If surveillance is everywhere, it is no longer a discrete process but the medium through which we move» (p. 87). Non si tratta più di plasmare il soggetto attraverso il timore del controllo, bensì di far coincidere lo spazio di sorveglianza con

l'ambiente stesso, in cui il soggetto opera come un agente fisico tra gli altri, di cui è possibile calcolare e valutare automaticamente i comportamenti per predirne (o prescriverne) le traiettorie.

L'audio viene a ricoprire, dunque, un ruolo strategico tanto nell'amministrazione quanto nella rappresentazione della smart city, la quale diviene la materializzazione del modello post-panottico di ambiente monitorato. Il suono, in virtù della sua ambientalità, del suo «essere intorno al soggetto», sembra essere l'emblema del nuovo paradigma. Con il suono, infatti, cade la barriera visibile tra spazi monitorati e spazi non-monitorati (Volcler, 2012). La sorveglianza ubiqua, di cui l'ascolto permanente è simbolo e rappresentazione, non è lo spettacolo panottico, né tanto meno lo «spionaggio *panacustico*» – che prevede ancora la presenza di un «punto d'ascolto» privilegiato di uno su molti (Szendy, 2008) – bensì il monitoraggio invisibile, pervasivo e desoggettivato che permette il funzionamento di algoritmi di previsione e intervento automatico.

5.1 Privacy

È evidente come ciò determini conseguenze per la questione della privacy (Lau, Zimmermann e Schaub, 2018). Da un lato, numerose preoccupazioni derivano da rappresentazioni fuorvianti che schiacciano l'ascolto operazionale su quello fenomenologico, non tenendo conto delle peculiarità di quello di poter scindere segnale e messaggio, ascoltare senza comprendere né memorizzare; come avviene per Amazon Alexa, in cui tutto ciò che non è riconosciuto come parola di attivazione è immediatamente cancellato (Salvador *et al.*, 2016). Dall'altro non è ben chiaro se e dove gli algoritmi di *audio sniffing* – in ascolto permanente di "parole chiave" da associare a un determinato profilo vocale a fini di pubblicità targetizzata e raccomandazioni – siano utilizzati (BBC, 2018; Edara, 2014)[9].

[9] Un interessante strumento per la tutela della privacy in ambito sonoro potrebbe essere il nuovo sistema di "anonimizzazione" della voce, che individuerebbe automaticamente i suoni vocali presenti accidentalmente nelle registrazioni ambientali, camuffandone sia il timbro che le parole (Cohen-Adria *et al*., 2019).

Ma significativi timori in merito alla privacy sono stati suscitati anche dalla recente notizia pubblicata da Bloomberg che Amazon avrebbe operatori umani all'ascolto delle interazioni vocali con Alexa (Day, Turner e Drozdiak, 2019)[10]. Questo episodio è rivelatore di un aspetto fondamentale, ovvero che per quanto le effettive modalità di estrazione ed elaborazione dati, di sorveglianza post-panottica, funzionino in maniera impersonale e desoggettivata, attraverso *clustering* trasversali di grandi quantitativi di dati, non vi sono rappresentazioni adeguate di questi sistemi, e le persone tendono ancora ad identificarli con la classica modalità disciplinare di spionaggio e controllo condotto da un individuo verso un altro individuo. Questo tipo di rappresentazione è più familiare e personale, poiché basata su umani specifici con le loro storie di vita, di cui la sorveglianza può conoscere intrusivamente i dettagli. Nell'ambito del progetto EAR-IT sono stati condotti, parallelamente agli studi di fattibilità della tecnologia, anche studi sulla "percezione" che i cittadini hanno del monitoraggio acustico. Tali studi confermerebbero una tendenza a pensare la privacy in termini di spionaggio: i cittadini sarebbero disposti a "farsi ascoltare" in cambio di un maggiore senso di sicurezza, a patto che i dati delle loro interazioni vocali e sonore con i dispositivi non vengano impiegati al di fuori dell'ambito strettamente securitario (Ståhlbröst, Sällström e Hollosi, 2014). Anche la dichiarazione di Florian Schaub dell'Università di Michigan pare non centrare il cambiamento che l'ascolto operazionale produce in ambito di privacy: «I think we've been conditioned to the [assumption] that these machines are just doing magic machine learning. But the fact is there is still manual processing involved» (Day, Turner e Drozdiak, 2019).

Per quanto sia vero che ci sono umani dietro le macchine, il punto è la mutata relazione gerarchica tra di essi. L'ascolto che quegli umani stanno facendo è interamente asservito alle finalità del *machine learning*: essi ripuliscono i dati, li mettono in ordine, li etichettano per ottimizzare i processi macchinici e accrescere l'accuratezza degli algoritmi di riconoscimento (un po' come avviene in SONYC). «We're creating, labeling, curating and analyzing vast quantities of speech on a daily basis» – recitano i portavoce di Amazon – «We only annotate an extremely small sample of Alexa voice recordings in order [to]

[10] Un caso simile è stato di recente documentato in merito all'assistente di Google (Verheyden *et al*., 2019).

improve the customer experience [to] train our speech recognition and natural language understanding systems» (Day, Turner e Drozdiak, 2019). In questo caso è l'ascolto umano a diventare operazionale, interamente circoscritto alla produzione di azioni strumentali il cui senso risiede nei principi di funzionamento dell'algoritmo.

Lo scandalo Amazon può essere visto come un esempio di rappresentazione ideologica: attribuisce agli operatori le tipiche facoltà legate all'ascolto fenomenologico, un ascolto che è immediatamente un memorizzare, un interpretare e un giudicare il nostro privato, ma non tiene conto, o meglio disconosce, le modalità principalmente macchiniche con cui quegli operatori, di fatto, utilizzano il loro senso dell'udito. Non colgono, dunque, lo spostamento fondamentale dell'umano da modello della macchina a *medium* per il funzionamento di questa che prima abbiamo definito, con Floridi, *computazione basata sull'uomo*.

La questione della privacy si gioca proprio su questo punto: la sorveglianza post-panottica e l'ascolto operazionale invitano, coerentemente con quello spostamento, a pensare ai dati non come qualcosa che le persone *posseggono*, ma come qualcosa che *forniscono*. Gli algoritmi non si rivolgono al soggetto incarnato ma al suo «doppio statistico» (Rouvroy e Berns, 2013, p. V), ai dati che egli fornisce ai sistemi computazionali e su cui essi operano i propri calcoli. Questo doppio statistico è un *profilo* che non esaurisce il soggetto, ma lo approssima quantitativamente e contribuisce alla costruzione di modelli algoritmici di previsione e di valutazione sempre più accurati e remunerativi.

La profilazione non è spionaggio, ma comprensione dei comportamenti di un'intera società – che dà ai profilatori, ai gestori di dati, notevoli vantaggi geopolitici e commerciali e permette un indirizzamento occulto delle scelte (Soro, 2019)[11]. Ciò esprime in maniera chiara come la questione della privacy si giochi in un equilibrio delicato tra immaginari e operazioni tecniche, dove i primi funzionano spesso da specchietto per le allodole, mentre le seconde insinuano gradualmente e surrettiziamente nuovi tipi di governamentalità.

[11] È significativo, al riguardo, che Amazon dedichi un'intera sezione dell'informativa sulla privacy di Alexa alle modalità di impiego dei dati per il *machine learning* dei suoi algoritmi di riconoscimento – sottolineando anche, probabilmente in risposta all'articolo di Bloomberg, che questa può prevedere l'analisi manuale oltre che automatica – ma neanche una riga all'uso dei dati per la profilazione degli utenti (Amazon, 2020).

6. Conclusioni

Dal quadro qui delineato emerge un generale cambio di paradigma di cui la smart city sarebbe allo stesso tempo causa ed effetto, attore e banco di prova. Decisiva per questo cambiamento è stata senz'altro l'introduzione di un particolare tipo di tecnologia, ovvero il *machine learning*, che fa dei dati digitali il motore dei propri sistemi di calcolo, i quali a loro volta possono tradursi in previsioni, classificazioni, decisioni, e in generale in potenti strumenti di *governance*.

Che i sistemi tecnologici odierni siano *data-driven* non è una necessità teorica, bensì un accidente storico (Zuboff, 2019); il loro potere di produrre servizi informatici impensabili fino a pochi decenni fa è indubbio, ma la loro epistemologia è tutt'altro che neutrale. La necessità di disporre di enormi *dataset* per far funzionare gli algoritmi, infatti, privilegia determinati campi su altri – e determinati attori economici – e distorce i saperi non quantitativi verso un approccio necessariamente numerico-statistico. Se, come recita il famoso articolo di Chris Anderson (2008), «con abbastanza dati i numeri parlano da soli», ovvero diventa possibile prevedere dei fenomeni senza, a rigore, "comprenderli" e conoscerne le cause, è il concetto stesso di comprensione a mutare. "Smart" sta ad indicare, quindi, una ben specifica forma di intelligenza: quella basata sull'analisi e la classificazione numerica di grossi quantitativi di dati, tra cui la macchina può rintracciare correlazioni e riconoscere *pattern* che permettano la predizione e valutazione automatica dei comportamenti. Se Chopplet (2018) riconduce l'idea di intelligenza "smart" al concetto di Dewey di «intelligenza votata all'azione», egli sottovaluta il fatto che la modalità "automatica" di quella corrispondenza non ha niente a che fare con l'intenzionalità di una coscienza fenomenologica, cioè che la "datità" su cui si fonda non è quella del dato per la coscienza, bensì quella molto diversa del dato estratto e trattato da macchine computazionali in virtù delle loro propensioni, che si realizza in un'azione automatica priva di quella conoscenza intellettuale prevista invece da Dewey.

In questo quadro si inserisce la pratica della sorveglianza nella sua riconfigurazione più recente, intesa non come dispositivo di governo dei soggetti, bensì come condizione tecno-

epistemica per il funzionamento degli algoritmi, ovvero come modalità di estrazione dati. Se il presupposto per il funzionamento dei sistemi *data-driven* è la misurazione, allora ogni fenomeno sociale deve poter essere ridotto a misura comune, e trattato di conseguenza come un evento fisico o naturale con strumenti matematico-statistici. L'enfasi si sposta dal soggetto al suo profilo numerico, *medium* per il funzionamento di sistemi algoritmici di valutazione e previsione.

La smart city non è l'evoluzione della città in senso razionalizzante, ma lo specifico dispositivo di governo prodotto dal nuovo paradigma della sorveglianza e della governamentalità algoritmica. E il suono, in virtù della sua invisibilità, incarna proprio questo ideale di sorveglianza, poiché fa venir meno il "punto di vista" del *panopticon* per diffondersi ambientalmente e ubiquamente.

Bibliografia

Amazon (2020). *Alexa and Echo devices are designed to protect your privacy*. Testo disponibile all'indirizzo web: https://www.amazon.com/Alexa-Privacy-Hub/b?ie=UTF8&node=19149155011 (07/12/2020).

Andrejevic M. (2019). *Automated Media*. New York: Routledge.

Anderson C. (2008). The end of theory. *Wired*, 23 giugno. Testo disponibile all'indirizzo web: https://www.wired.com/2008/06/pb-theory/ (04/09/2020).

Barthes R. (2001). Ascolto. In Barthes R., *Saggi critici. Vol. III: L'ovvio e l'ottuso*. Torino: Einaudi.

BBC (2018). *Amazon patents "voice-sniffing" algorithms*. Testo disponibile all'indirizzo web: https://www.bbc.com/news/technology-43725708 (04/09/2020).

Bello J.P., Silva C., Nov O., Dubois R. L., Arora A., Salamon J., Mydlarz C., Doraiswamy H. (2019). SONYC: A system for the Monitoring, Analysis and Mitigation of Urban Noise Pollution. *Communications of the ACM*, 62, 2: 68. DOI: 10.1145/3224204.

Bonnet F. (2016). *The Order of Sounds. A Sonorous Achipelago*. Falmouth: Urbanomic.

Borrelli D. (2000). *Il filo dei discorsi. Teoria e storia sociale del telefono.* Milano: Luca Sossella Editore.

Borrelli D. (2015). *Contro l'ideologia della valutazione. L'Anvur e l'arte della rottamazione dell'Università.* Sesto San Giovanni: Jouvence.

Chopplet M. (2018). Smart City: quelle intelligence pour quelle action? Le concept de John Dewey scalpel de la ville intelligente. *Quaderni*, 96, 2: 71. DOI: 10.4000/quaderni.1179.

Cohen-Hadria A., Cartwright M., McFee B., Bello J.P. (2019). Voice Anonymization in Urban Sound Recordings. *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. Pittsburgh, PA: IEEE. DOI: 10.1109/MLSP.2019.8918913.

Commissione Europea (2014). *EAR-IT: Using sound to picture the world in a new way*. Testo disponibile all'indirizzo web: https://ec.europa.eu/digital-single-market/en/news/ear-it-using-sound-picture-world-new-way (04/09/2020).

CORDIS (2017). *Experimenting Acoustics in Real Environments using Innovative Test-beds*. Testo disponibile all'indirizzo web: https://cordis.europa.eu/project/id/318381 (04/09/2020).

Cox T. (2018). *Now You're Talking: Human Conversation from the Neanderthals to Artificial Intelligence*. Berkeley: Counterpoint.

Day M., Turner G., Drozdiak N. (2019). *Amazon workers are listening to what you tell Alexa*. Testo disponibile all'indirizzo web: https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio (04/09/2020).

Delale S. (2019). *How a city can become smarter with voice*. Testo disponibile all'indirizzo web: https://www.voicesummit.ai/blog/how-a-city-can-become-smarter-with-voice (04/09/2020).

Dolar M. (2014). *La voce del padrone. Una teoria della voce tra arte, politica e psicoanalisi*. Napoli: Orthotes.

Doran D., Gokhale S., Dagnino A. (2013). Human Sensing for Smart Cities. ASONAM '13: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks

Analysis and Mining. Association for Computing Machinery: New York. DOI:10.1145/2492517.2500240.

Edara K.K. (2014). *Key word determinations from voice data*. Brevetto n. US8798996B1. Testo disponibile all'indirizzo web: https://patentimages.storage.googleapis.com/bd/ed/2b/c4c67cc5a9f1ab/US8798995.pdf (04/09/2020).

Ernst W. (2016). *Sonic Time Machines*. Amsterdam: University of Amsterdam Press.

Euronews (2014). *In the smart city of Santander the walls have ears.* Testo disponibile all'indirizzo web: https://www.euronews.com/2014/10/13/in-the-smart-city-of-santander-the-walls-have-ears (04/09/2020).

Floridi L. (2017). *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina.

Fraunhofer IDMT (2014). *SonicSentinel: an intelligent sound monitor for care homes*. Testo disponibile all'indirizzo web: https://www.idmt.fraunhofer.de/content/dam/idmt/documents/HSA/SonicSentinel_Fraunhofer_IDMT_en.pdf (04/09/2020).

Fraunhofer IDMT (2019). *Smart acoustic sensors for minimum machine downtime*. Testo disponibile all'indirizzo web: (https://www.idmt.fraunhofer.de/en/Press_and_Media/press_releases/2019/smart-acoustic-sensors-for-minimum-machine-downtime.html (04/09/2020).

Garnier J.P. (2019). *Smart City. La "città radiosa" nell'era digitale.* Torino: Nautilus.

Goodfellow I., Bengio Y., Courville A. (2016). *Deep Learning.* Cambridge: MIT Press.

Greenfield A. (2006). *Everyware: The dawning age of ubiquitous computing.* Boston: New Riders.

Habermas J. (1981). *Teoria dell'agire comunicativo*. Bologna: il Mulino.

Hollosi D., Nagy G., Rodigast R., Goetze S., Cousin P. (2013). Enhancing Wireless Sensor Networks with Acoustic Sensing Technology: Use Cases, Applications & Experiments. *IEEE International Conference on Internet of Things (iThings), Beijing, China*. DOI: 10.1109/greencom-ithings-cpscom.2013.75.

Iaconesi S., Persico S. (2015). Il Terzo Infoscape. Dati, informazioni e saperi nella città e nuovi paradigmi di interazione urbana. In Arcagni S., a cura di, *I media digitali e*

l'interazione uomo-macchina. Ariccia: Aracne Editore.

Kelly B., Hollosi D., Cousin P., Leal S., Iglar B., Cavallaro A. (2014). Application of Acoustic Sensing Technology for Improving Building Energy Efficiency. *Procedia Computer Science*, 32: 661. DOI: 10.1016/j.procs.2014.05.474.

Kitchin R. (2014). The real-time city? Big data and smart urbanism. *GeoJurnal*, 79: 1. DOI: 10.1007/s10708-013-9516-8.

Lau J., Zimmerman B., Schaub F. (2018). Alexa, are you listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proceedings of the ACM on Human-Computer Interaction*, 2, CSCW, Article 102. DOI: 10.1145/3274371.

Li X. (2017). *Divination Engines: A media history of text prediction and speech recognition*. New York: ProQuest Dissertations Publishing.

Marcuse H. (1999). *L'uomo a una dimensione*. Torino: Einaudi.

Marx K. (2004). *Manoscritti economico-filosofici del 1844*. Torino: Einaudi.

Mauss M. (2017). *Le tecniche del corpo.* Roma: ETS.

Mills M. (2012). Media and Prosthesis. The Vocoder, the Artificial Larynx, and the History of Signal Processing. *Qui Parle. Critical Humanities and Social Sciences*, 21, 1: 107. DOI: 10.5250/quiparle.21.1.0107.

Mohler G. (2015). Predictive Policing: George Mohler Interview. *Data Science Weekly*. Testo disponibile all'indirizzo web: https://www.datascienceweekly.org/data-scientist-interviews/predictive-policing-george-mohler-interview (04/09/2020).

Morozov E. (2018). Parte I. In Morozov E., Bria F., *Ripensare la smart city*. Torino: Codice.

Mumford L. (2002). *La città nella storia*. Milano: Bompiani.

Nirjon S., Srinivasan R., Sookoor T. (2017). Smart Audio Sensing-Based HVAC Monitoring. In Song H., Srinivasan R., Sookoor T., Jeschke S., a cura di, *Smart Cities: Foundations, Principles and Applications.* Hoboken: Wiley.

Norman D.A. (2010). Natural user interfaces are not natural. *Interactions*, 17, 3: 6. DOI: 10.1145/1744161.1744163.

Peters J.D. (2004). Helmholtz, Edison and sound history. In Rabinovitz L., A. Geil, a cura di, *Memory Bytes. History, Technology and Digital Culture*. Durham e Londra: Duke

University Press.

Pham C., Cousin P. (2013). Streaming the Sound of Smart Cities: Experimentations on the SmartSantander Test-bed. In IEEE, *IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. Pechino: IEEE. DOI: 10.1109/GreenCom-iThings-CPSCom.2013.114.

Pieraccini R. (2012). *The voice in the machine. Building computers that understand speech.* Cambridge: MIT Press.

Roe D.B., Wilpon J. G., a cura di (1994). *Voice communication between humans and machines*. Washington D.C.: National Academy of sciences. DOI: 10.17226/2308.

Rouvroy A., Berns T. (2013). Algorithmic governmentality and prospects of emancipation. *Résaux*, 177: 163. DOI 10.3917/res.177.0163.

Sahib U. (2020). Smart Dubai: Sensing Dubai Smart City for Smart Environment Management. In Vinod Kumar T., ed., *Smart Environment for Smart Cities. Advances in 21st Century Human Settlements*. Singapore: Springer.

Salvador S.W., Lilly J.P., Weber F.V., Adams J.P., Thomas R.P. (2016). *Wake word evaluation*. Brevetto n. US9275637B1. Testo disponibile all'indirizzo web: https://patentimages.storage.googleapis.com/4b/e0/d6/0a568b8cd78657/US9275637.pdf (04/09/2020).

Schaeffer P. (2017). *Treatise on Musical Objects*. Oakland: University of California Press.

Simonofski A., Asensio E.S., De Smedt J., Snoeck M. (2019). Hearing the Voice of Citizens in Smart City Design: The CitiVoice Framework. *Business & Information Systems Engineering*, 61: 665. DOI: 10.1007/s12599-018-0547-z.

Soro A. (2019). *Democrazia e potere dei dati. Libertà, algoritmi, umanesimo digitale*. Milano: Baldini&Castoldi.

Srnicek N. (2017). *Capitalismo digitale. Google, Facebook, Amazon e la nuova economia del web*. Roma: Luiss University Press.

Ståhlbröst A., Sällström A., Hollosi D. (2014). Audio monitoring in smart cities. An information privacy perspective. In Kommers P., Isaías P., a cura di, *Proceedings of 12th IADIS International Conference e-Society*. Testo disponibile all'indirizzo web:

http://www.diva-portal.org/smash/get/diva2:1005773/FULLTEXT01.pdf (22/10/2020).

Sterne J. (2003). *The Audible Past. Cultural Origins of Sound Reproduction.* Durham e Londra: Duke University Press.

Szendy P. (2008). *Intercettare. Estetica dello spionaggio.* Milano: Isbn Edizioni.

Tarizzo D. (2013). Dalla biopolitica all'etopolitica: Foucault e noi. *Nòema*, 4, 1: 43. DOI: 10.13130/2239-5474/2871.

Unione Europa (2011). *La città del futuro. Sfide, idee, anticipazioni*. Testo disponibile all'indirizzo web: https://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/citiesoftomorrow/citiesoftomorrow_summary_it.pdf (04/09/2020).

Verheyden T., Baert D., Van Hee L., Van Den Heuvel R. (2019). *Google employees are eavesdropping, even in your living room, VRT NWS has discovered*. Testo disponibile all'indirizzo web: https://www.vrt.be/vrtnws/en/2019/07/10/google-employees-are-eavesdropping-even-in-flemish-living-rooms (04/09/2020).

Voegelin S. (2010). *Listening to Noise and Silence. Towards a Philosophy of Sound Art.* New York: Continuum.

Volcler J. (2012). *Il suono come arma. Gli usi militari e polizieschi dell'ambiente sonoro*. Roma: DeriveApprodi.

Wang A. (2003). An Industrial Strength Audio Search Algorithm. *Proceedings of ISMIR 2003, 4th International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, Maryland, USA, October 27-30. Testo disponibile al sito web: https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf (04/09/2020).

Zizek S. (2012). *Il resto indivisibile. Su Schelling e questioni correlate*. Napoli: Orthotes.

Zuboff S. (2019). *Il capitalismo della sorveglianza.* Roma: Luiss University Press.