Productive Sounds

Touch-Tone Dialing, the Rise of the Call Center Industry and the Politics of Virtual Voice Assistants

Axel Volmar

The growing dissemination of virtual voice assistants in smartphones, smart speakers and vehicle onboard systems, such as Amazon's Alexa, Apple's Siri, Google's Assistant, Microsoft's Cortana or Samsung's Viv, represent a democratization of artificial intelligence by sheer mass exposure.[1] Voice assistants, generally referred to as intelligent virtual assistants (IVAs) or intelligent personal assistants (IPAs), belong to a class of software agents that can answer queries and perform tasks for users based on verbal commands and inquiries when equipped with a voice user interface (VUI). Tech corporations promote their voice-centered smart assistants as pinnacles of contemporary artificial intelligence and as new forms of seamless cooperation between man and machine, built to offer more intuitive ways of controlling and navigating digitally networked and cloud-based technology. The imminent ubiquity of conversational AI, however, raises a number of fundamental questions regarding algorithmic control as well as the nature and history of sound-based human–machine interaction. How are these emergent forms of voice-based cooperation structured and how does voice control change our relationship with and critical assessment of software technology? What ramifications result from AI technologies being based largely on cloud computing and thus from user data being sent to cloud servers to be processed?

Given the black-box character of most commercially available AI technologies, it is naturally rather difficult to obtain detailed information about how the AI algorithms of particular voice assistants exactly function. However, it is not necessary to understand how they work algorithmically in every detail to understand their politics; it is sufficient to study what they are used for and how they are marketed to different stakeholders and actors. I therefore conceptualize intelligent personal assistants—on mobile phones, operating systems, and especially smart

1 A recent report by market analyst firm Canalys (2019) predicts that the worldwide smart speaker install base is set to grow 82.4 per cent from 114 million sold units to over 200 million by the end of 2019.

speakers—as *platforms* in the sense of media scholar Tarleton Gillespie. In his well-received paper, Gillespie argues that the politics of platforms can be traced by examining how

> online content providers such as YouTube are carefully positioning themselves to users, clients, advertisers and policymakers, making strategic claims for what they do and do not do, and how their place in the information landscape should be understood. One term in particular, 'platform', reveals the contours of this discursive work. (Gillespie 2010: 347)

Similarly, I will focus in this paper less on the inner workings of the machines themselves than on the various relations of voice interfaces to their immediate surrounding environment and on the purposes they serve for different actors, such as users, call center agents, businesses, major tech corporations, and surveillance states. However, I will take a considerable historical detour in the effort to ground conversational AI in a broader history of sound- and voice-based human–machine interaction and to emphasize continuities and caesuras between contemporary voice assistants and previous sound- and voice-based user interfaces for networked services. Another reason for this approach is that despite the current hype around voice assistants, auditory and speech-based human–machine interfaces are far from being recent developments. Ever since the psychologically troubled board computer HAL from Stanley Kubrick's *2001: A Space Odyssey* (1968), speech interfaces for human–computer interaction have had a permanent place in the cultural imaginary of industrialized societies.

Although sophisticated artificial intelligence systems like HAL still remain science fiction, sound and speech indeed represent one of the oldest interfaces for interacting with remote systems. However, early applications did not emerge in the computer industry but in the telecommunications sector. Shortly after the release of *2001*, AT&T promoted its Touch-Tone telephones for queries in digital-inquiry/voice-answer (DIVA) systems, which allowed for information retrieval in the form of computer-controlled voice messages through and triggered by Touch-Tone commands. Telephonic practices of interacting with distributed services via sound and speech date back to even the 1940s and 1950s, before they were further developed in the growing call center industry. Contemporary practices of speaking to machines therefore reinterpret forgotten or discarded user experiences connected to the telephone. To this effect, I second media scholar Jonathan Sterne's (2012) emphasis on the centrality of telephony and sound technologies to the history of digitality:

> Telephony is often considered anaesthetic matter in comparison with the usual, more aestheticized subjects of twentieth-century media history such as cinema,

> television, sound recording, radio, print, and computers. But telephony and the peculiar characteristics of its infrastructure are central to the sound of most audio technologies over the past 130-odd years. The institutional and technical protocols of telephony also helped frame the definitions of communication that we still use, the basic idea of information that subtends the whole swath of "algorithmic culture" from packet switching to dvds and games, and the protocols and routines of digital technologies we use every day. (2-3)

While Sterne used the history of the telephone system, and especially developments in signal compression methods and perceptual coding to unpack the mp3 format as a "cultural artifact" (Sterne 2006), I discuss speech-related artificial intelligence applications against the backdrop of a longer history of remote telephone services and processes of (semi-)automation in the telecommunications and customer service industry, with particular attention to call centers. Automation has been a driving force, if not the condition of possibility, of call centers from the very beginning. Most of these attempts are based on what I want to call *productive sounds*, i.e., sounds that serve specific purposes within a (semi-)automated system or even literally perform work, such as triggering switching or algorithmic processes.[2] Productive sounds such as Touch-Tone signals, hold music and recorded voice messages lie at the center of a transformational process in which telephone companies aimed to extend the telephone system from a special-purpose application for voice transmission into a general-purpose information network (cf. Lipartito 2003). Taking the form of synthesized voices in conversational AI and digital personal assistants, sounds became productive as special-purpose substitutes for general-purpose manual tasks previously performed by computer users.

In media theoretical terms, we can understand this transition by conceptualizing productive sound media not as media of *communication* but, in the words of German media theorist Erhard Schüttpelz, as potentially powerful media of *cooperation* (Schüttpelz 2017: 14; cf. Volmar 2017). For instance, to speak of the telephone as a cooperative medium means to conceive it not as a mere conversational medium but as a more universal means to facilitate logistical, bureaucratic, problem-solving, and other quotidian personal tasks of work-related "infrastructuring" (Star/Bowker 2002). At a time when we casually associate such logistical tasks with the internet, online platforms, mobile apps or smart speakers, it seems worth a reminder that the underlying narrative of inter-networked information services is actually older than the internet itself and that it once was deeply entangled with

2 While I use the term "productive sounds" in this specific sense, I take the general notion from Alix Hui and Joeri Bruyninckx who introduced the term at their workshop "Productive Sounds in Everyday Spaces: Sounds at Work in Science, Art, and Industry, 1920–Present" at the Max Planck Institute for the History of Science on April 27-28, 2018.

circuit-switched telecommunications infrastructure. I argue that voice-centered AI applications in call centers (now usually referred to as 'contact centers') and domestic environments can be regarded as a current escalation within the history of cooperative sound media and the various attempts to automate the practices that revolve around them.

Cooperation always entails practices performed by and between different actors and groups. To highlight developments in cooperative practices within the history of voice automation, I pay particular attention to forms of phone- and voice-related work and labor practices. While scholars in the history of media and technology have extensively studied the work of telephone operators (e.g., Green 1995; Lipartito 1994), I follow media and sound scholar Sumanth Gopinath's work on the ringtone industry (Gopinath 2013) by focusing on the significance of sonic and telephonic labor within the infrastructural frameworks of the customer service industry to trace the formation of networked, speech-based human–machine interactions. To this end I examine how changing distributions and delegations of work between call center agents and customers as well as between humans and machines constitute infrastructures of *tele cooperation*, parts of which we also find in current digital assistants.

In section 1, I take a step back to revisit the ramifications of AT&T's introduction of the push-button telephone in the early 1960s. Initially sought to replace operators by further automating the initiation and switching of telephone calls, push-button telephones featured the new dialing method of dual-tone multi-frequency (DTMF) signaling, which operated on the basis of "in-band", i.e., audible control signals—the dial tones we still hear in landline and mobile phones when pushing buttons on the keypad. Sometimes the tones are even simulated on smartphones, for instance within messenger apps. I argue that while multi-frequency signaling rendered telephone switching more automatic and efficient, it also led to practices of delegating and outsourcing phone work from operators to both automatic systems and customers.

More importantly, MF signaling enabled the transmission of sonically coded alpha-numerical information over the telephone network and thus formed a fundamental condition of possibility for the emergence of automatic phone-based information systems in modern call centers. In section 2, I recall some of these technological innovations, especially automatic call distributors (ACDs) and interactive voice response systems (IVRs), both of which were foundational for the rise of the call center industry. I then examine how these contributed to the semi-automation of telephone calls and the further redistribution of voice and sound work by breaking down telephone conversations into common inquiries and sequences and how both call center agents and callers had to adjust themselves to these standardized "boundary objects" (Star/Griesemer 1989) in order to make the automated systems work.

In section 3, then, I show how artificial intelligence entered the stage in the contact center, as it had come to be called, in the form of speech recognition, understanding, and synthesis. I argue that decades of semi-automating phone calls and adjusting agents and customers to automated systems made the contact center particularly receptive to artificial intelligence technology within the industry. The implementation of conversational AI is based on a similar logic as IVRs, as it mainly breaks down phone conversations into a limited number of categories or entities, such as certain key words or presumed emotional states. The same logics are present in contemporary voice assistants for the home. By situating contemporary voice assistants within the broader history of semi-automation and cooperative telephonic practices based on productive sounds and voice work in the call center industry, I ultimately seek to expand existing histories of the internet and digital culture (e.g., Haigh et al. 2015) by considering the evolution of telephone-based telecommunications as an important area for the conception, testing, and mainstreaming of digitally networked media and cooperative practices.

1. Push-button Telephones and Touch-Tone Dialing: Innovation in General-Purpose Infrastructural Technologies

In the first half of the twentieth century, the handling of telephone calls in the Bell System largely remained in the hands of human telephone operators, even though a number of solutions for automatic switching, such as the Strowger switch, were at hand. Whereas technical issues and a reluctance of Bell System managers to license external patents on automatic switching formed the major reasons for clinging to manual switching (Green 1995; Lipartito 1994), opponents of automatic switching argued that establishing the connection represented a form of technical work that should be offered as part of the telephone service and hence done by operators. Harris F. Hopkins, the author of an article in the *Bell Laboratories Record*, put it this way: "Oppositionists felt that automatic switching was wrong from the customer's viewpoint. 'The public will not tolerate doing its own operating,' they said" (Hopkins 1960: 83). After the Second World War, however, rotary-dial telephones to automate the initiation of local phone calls became increasingly common. This transition to self-operating shows that the central logic of automation extended beyond the simple substitution of work by machines to the delegation or redistribution of work in general, in this case from service providers to their customers. The outsourcing of labor to both machines and customers in order to save labor cost, which forms a signature of today's digital culture, was already an economic driving force in the postwar telecommunications sector.

On November 18, 1963, Bell introduced yet another innovation in dialing automation: the push-button telephone, which featured not just a different way of

manual dialing but an entirely new way of creating dialing signals. Dialing on a rotary phone produced a train of electrical impulses, the number of which corresponded to the indicated digit on the rotary dial. Pressing a button on a push-button telephone, however, created a distinct pair of two audible sine tones generated by electronic oscillators. This so-called dual-tone multi-frequency (DTMF) dialing method was based on a four-by-four frequency scheme proposed by L. A. Meacham of Bell's Station Development Department, although initially only seven frequencies (four in the low end of the spectrum and three in the higher range of the spectrum) would generate ten unique pairs of tones (Meacham et al. 1958).[3]

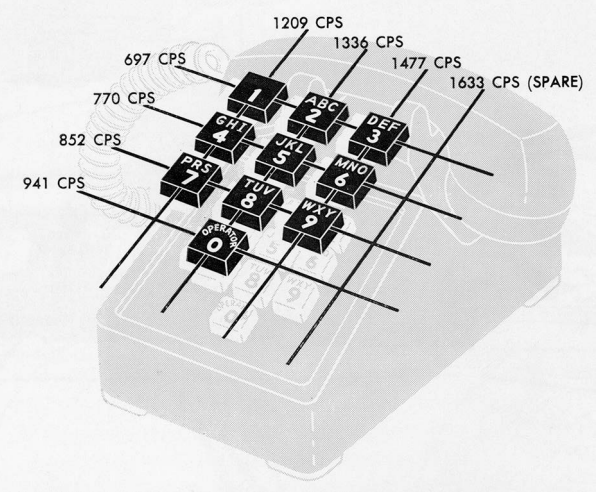

Fig. 1: Four-by-four frequency scheme for the generation of DTMF dialing signals. Image source: Noweck 1961: 314. Courtesy of AT&T Archives and History Center.

The method made use of state-of-the-art solid-state technology and was grounded in a number of field trials conducted between 1948 and 1960 (Dahlbom et al. 1949; Hopkins 1960). When dialing a number, the dual-tones provided a helpful acoustic feedback for the caller. Nevertheless, the sounds were not addressed to human ears to hear in the first place but to electronic filter banks, which were installed at the local switching stations, the so-called "call centers," for decoding. To prevent

3 The pairing of tones followed a simple rule of construction. Each vertical column has a different tone in the low frequency range assigned (FA = 697 Hz, FD = 770 Hz, FC = 852 Hz und FD = 941 Hz), while each horizontal row has a different higher frequency tone assigned (FE = 1209 Hz, FF = 1336 Hz, FG = 1447 Hz und FH = 1633 Hz). This way, each key is assigned to a different combination of a high and a low frequency tone. The necessary hardware comprised a keypad encoder and tone generator.

spoken language, noises and other sounds from interfering with the transmission of DTMF tones, the microphone was disconnected when pressing down a button. Further, because the dialing signals were audible "in-band" frequencies, the Bell technicians chose combinations of frequencies that were unlikely to occur in everyday life so as to prevent false positives and false negatives from occurring in the receivers of the switching equipment: "The frequencies that are used minimize interference from harmonics. This permits instantaneous limiting in both frequency bands, and satisfactorily guards against possible voice interference" (Hopkins 1960: 86). If you ever wondered why push-button tones sound more like the otherworldly noises of electronic music than the harmonious sounds of musical instruments, this is why.

At the 1964 World's Fair in New York, Bell presented DTMF signaling to the public under the brand name Touch-Tone. By means of Touch-Tone calling, subscribers were enabled to initiate, for the first time, long-distance calls directly without the need of a human operator as an intermediary. The introduction of the push-button telephone was therefore closely related to the more or less simultaneous introduction of electronic switching systems (ESSs) to the central switching stations. ESSs were based on digital "stored program control" (SPC), an automated and computerized method of monitoring telephone switching developed around 1954 by Bell Labs mathematician Erna Schneider Hoover (Harr et al. 1964). Electronic switching proved to be more stable and reliable than mechanical methods and eliminated almost entirely the need for human operators. Since tone-based dialing was vital for the introduction of digital switching, the use of sound was also part of a foundational step in the history of digitization. The main advantage of DMTF dialing was the fact that tones could be both generated and detected much faster than the pulse signals generated by rotary phones. The increased speed was particularly helpful for long-distance calls and calls to individual extensions, for instance within larger organizations, since this could greatly increase the number of digits to dial and hence demanded time and patience on the part of the caller. While Bell promoted Touch-Tone dialing to its customers as a more convenient way of initiating calls, the method was particularly tailored to unburden the switching centers, where the old step-by-step switches that could become serious bottlenecks in the connection process, especially during peak calling times. With Touch-Tone signaling, switching centers were able to handle many more calls within a much smaller time span.

The adoption of in-band signaling, however, was not intended to improve the dialing process and the handling of calls alone but to enable new ways of interacting with electronic, and possibly digital, systems connected to the telephone network. As Hopkins (1960) points out, Bell had confidence in offering this "possible future service" because Touch-Tone dialing would provide "the customer with a potential (slow-speed) data transmitter" (87). The first widely distributed

push-button telephone was Western Electric's Model 1500, which came with 10 buttons corresponding to the digits 0 through 9 (see fig. 1). On later models, buttons with the now ubiquitous number (#) and star (*) signs were added to enable and control the transmission of symbolic data. Transmogrified into a potential remote control or terminal device, the telephone receiver could be used to provide alpha-numerical information, such as credit card numbers, or place commands, such as vertical service codes (VSCs). VSCs are sequences of digits in combination with the signals star (*) and, less frequently, number sign (#). Dialled on a telephone keypad or rotary dial, a VSC could be used to enable or disable certain telephony service features, such as call hold, call forwarding, continuous redial or call blocking. The term "vertical" refers to commands pointing to higher-level instructions within the local telephone infrastructure rather than regular telephone numbers, which point out "horizontally" to another geographic location or switching center. AT&T began to introduce VSCs under the name "Custom Local Area Signaling Services" (CLASS or LASS) codes to subscribers in the 1960s and 1970s. With Touch-Tone, sound thus became an acoustic interface for interactions with automated electronic and digital systems.

Seen from the perspective of speech act theory (Austin 1975), the DTMF tones can be conceptualized as "sonic acts" or "sound acts," i.e., as sounds that not only *represent* something or contain information but also *act* and have consequences. As audible control signals, designed to communicate with automated electronic systems over the network, DTMF tones literally became *productive* sounds within the telephone system as they triggered switches, transmitted information and remote-controlled automatic processes. It was on the basis of productive sounds, then, that Bell engineers aimed to prepare the telephone system for the information age. Or put another way, Bell engineers realized that a technology conceived for optimizing their own infrastructure could also be used to develop and offer new information services to both their business and domestic customers. In regard to practice, the growing habit of dialing telephone numbers and using other services, such as VSCs, contributed to training subscribers to perform different forms of *data work*. As noted above, Touch-Tone dialing enabled end-to-end signaling, the transmission of control signals not only to the nearest switching center but also to switching systems anywhere in the network. Therefore, the DTMF method needs to be regarded as an infrastructural medium that played a fundamental role in the transformation of the telephone from a special-purpose technology for talking over distances to a general-purpose technology for speech, data transmission and remote control.

Fig. 2: Different potential applications for banking, retail, or domestic use interconnect customers and digital systems via Touch-Tone telephones. The original caption reads: "Many businesses are using the double-duty TOUCH-TONE® telephone and a computer to speed customer services and develop new ones as well. Banks use the Touch-Tone phone in an information retrieval system known as DIVA (for Digital Inquiry-Voice Answer). With this system, for example, a teller can query the bank 's central computer for a customer's up-to-elate balance before cashing a check (upper left). He dials the computer, taps a few buttons to identify the account number (or, if his phone is a card-dialer model as shown, inserts a DIVA account card) and the code for current balance. The computer responds with a voice answer. Data systems using the Touch-Tone telephone are being used by clerks in retail stores as well. As shown (upper right), the clerk telephones a computer to record each sale she makes. In this case, she sends the account number (for credit sales), the price, merchandise code, and her own clerk number. Billing and accounting are then handled automatically. Eventually, even a house wife (left [image not reproduced here]) may use the Touch-Tone telephone to "shop by phone," pay bills, or check her bank balance." Image source: Soderberg 1969: 203. Courtesy of AT&T Archives and History Center.

2. Speaking to Machines, Speaking in Code: The Rise of the Call Center Industry and the Semi-automation of Phone Conversations

AT&T began to offer new custom calling services based on Touch-Tone dialing in the mid 1960s. These featured new functionalities, such as call waiting, call forwarding, and three-way service or conference calls. Moreover, automatic data collection and information retrieval systems, such as the digital-inquiry/voice-answer (DIVA) system (see the textbox in fig. 2), were sought to bring new forms of distributed cooperation to the business world and domestic subscribers. Bell engineers envisioned diverse workflows of quotidian "infrastructuring" (Star/

Bowker 2002) in a number of different domains, such as banking, retail, and personal use (see fig. 2). For J. H. Soderberg, who summarized some of the potential commercial applications of Touch-Tone-based services in 1969, the switched telephone network pointed the way into the digital future of networked devices and distributed services:

> The possibilities for using the Touch-Tone telephone for control purposes are virtually unlimited. Not only can the Touch-Tone telephone bring the computer revolution into every living room or office across the nation, but it can perform many other simpler control functions. It is even conceivable that future systems will permit you to turn on your home air conditioner so that your home will be comfortable when you return from a trip, or let you "shop by phone"—merely by pushing a few buttons on your telephone. The result could be a dramatic simplification of everyday tasks. (Soderberg 1969: 203)

As Soderberg's vision shows, Bell engineers and marketers had surprisingly clear ideas about the potential of digitally networked, semi-automated services in telephone banking, distributed accounting, home shopping and smart home applications. Not least due to antitrust laws, which banned cross-subsidizing "enhanced" telecommunications services and largely prevented AT&T from venturing into computer businesses, many of these possible applications remained for more than another decade just that, a technological potential and good publicity for the Touch-Tone service. It took well until the 1980s before push button phones reached a considerable saturation.[4] But watch any Hollywood film from the time featuring 1980s yuppie culture and you will see Touch-Tone services everywhere and realize: the telephone system was the internet of services before the internet of services.

Touch-Tone-based services, however, proved tremendously successful in the customer service sector and were deeply connected to the rise of call centers. In the late 1950s and early 1960s, call centers began to form in the offices of telephone companies for their own customer and operator support. Two technical innovations fostered the spread of premise-based call centers. First, the introduction of private automated branch exchanges (PABX), later also referred to as private automated business exchanges, allowed automatic routing to an extension number in a larger organization and hence replaced the work of phone receptionists or attendants (see Bodin 2002: 20). Shortly after, automatic call distributors (ACD) extended PABX capability to collect incoming calls—for instance, to the central

4 The technology was still considered a "premium" feature until well into the 1990s, when personal computers connected to the internet via modems began to challenge the use of the telephone as the go-to interface for interacting with distributed online-services.

office of an organization—and route them to a group of customer service agents.[5] In case all agents were busy, the ACD placed the incoming call in a waiting line until an agent became available. The functionality of ACDs is based on sophisticated algorithms, such as Erlang calculations, for predicting how many agents are needed and how to best queue and assign large numbers of simultaneous calls. ACDs can therefore be seen as the foundation of call centers and represent the first kind of artificial intelligence (in the larger sense of the word), because they introduce automatic decision making to the management of calls. However, ACDs are not artificial intelligence in the narrow sense of the term but rather "conditional call routing solutions, based on if-then conditions, or rules pre-defined by the organization" (Stanley 2018). Nevertheless, ACDs assure to this day that callers are answered as quickly as possible and that the time of all agents is used evenly and effectively.

Both PABXs and ACDs reduced the need for human operators and receptionists in central telephone offices and even rendered their work entirely obsolete. Moreover, sophisticated ACDs provided reports on various aspects of the call transaction (Bodin 2002: 22-23). Automatic call distributors proved particularly valuable for organizations that faced large call volumes. However, automatic in-house routing had the obvious disadvantage, due to algorithmic procedures, of not allowing callers to contact an agent directly. Since callers were unlikely to get assigned to the same agent twice, it prevented them from forming relationships with particular agents and hence resulted in a much less personal calling experience. AT&T's introduction of toll-free 1-800 numbers in 1967 basically established automatic call distribution on a nationwide scale—the service would first redirect calls to a national or local call center, where on-premise ACDs would further route the call to available agents.[6] Toll-free numbers led to an unprecedented increase in customer service call volume and cemented the anonymous user experience as a *de facto* standard. ACDs became the foundation of large-scale, decentralized and geographically distributed call centers. Among the early ACD solutions that proved economically successful, the US-manufacturer Rockwell is one of the most credited. The company's Galaxy ACD, as the device was called, enabled Continental Airlines to start offering phone-based flight reservation in 1973.

In the 1970s, the potential of DTMF signaling was recognized by manufacturers of call center equipment. So-called interactive voice response (IVR) systems automated not only the routing of calls but also specific parts of the actual phone conversations themselves. ACDs could play welcome messages, but they

5 The job of automatic call distributors, or ACDs, is to filter, order and assign incoming calls to the best available agent.

6 The inventor of the toll-free number once stated that all he had invented was in fact a pointer in a digital directory.

featured no further functionality other than putting the caller on hold. In IVRs, prerecorded messages would inquire about the caller's needs, acoustically guide them through a menu structure and present them with choices for different services, which the caller would then be able to select by pressing the corresponding buttons on a push-button phone. Typically, these systems were semi-automated human–machine systems with IVRs at the front end and human agents who took over at predefined points or whenever an automated system would come up against limits. The division of labor between humans and machines was achieved by breaking down phone conversations into parts with greater or lesser degrees of redundancy and automating the former. Fixed sets of categories and options addressed most customer queries, delivered through prerecorded messages that caller callers could respond to using DTMF tones. We can therefore regard the relation between the customer and a respective organization, which unfolds within an IVR system, as what Susan Leigh Star and James R. Griesemer have called a "cooperation without consensus" based on a common techno-conversational "boundary object" (1989).

The self-service functionality of IVRs allowed for substituting, at least in part, not only operator work but also the actual voice and transactional work performed by customer service agents. Other than the obvious saving of labor cost, automatic call center systems had the advantage of enabling expanded service hours. The flipside, however, was that since callers were not even talking to human agents anymore—at least not until the system connected them to one—IVRs rendered the phone experience even more anonymous than the seemingly random selection process done by automatic call distributors. Over the years, vendors added voice recognition to Touch-Tone as an alternative input language. The primary goal of introducing voice control had been to extend IVR services to owners of rotary-dial telephones but the result was that with voice recognition, whoever preferred speaking to typing was now able to interact with the IVR system via spoken language. This is the point where AI techniques first enter the stage.

Most of these circuit-switched telephonic systems have since been replaced by packet-switched, IP-based technology. Their story is therefore, at least to some extent, also an archaeology or reconstruction of media-cultural visions of a semi-automated future, consisting of human operators and interactive systems. They also refer to a hybrid future of cooperative systems that were both analog and digital at the same time. George Lucas' first feature film, *THX 1138* (1971), is exemplary of the future visions in this period of telephonic information networks. Lucas paints a picture of a futuristic underground society permeated by communication and surveillance technologies, reminiscent of George Orwell's novel *1984*. He thereby extrapolates contemporaneous advancements in touch-button telephones and semi-automatic systems into a dystopia of total audiovisual mediatization and surveillance. The impression of the omnipresence of media-technological media-

tion and observation is further reinforced by frequently staging technically mediated communication situations in the form of telephone and intercom conversations, tape announcements, video transmissions, and CCTV images.

As a response to its cultural moment, *THX 1138* forms an artistic reflection on the then-incipient transformation of acoustic media into what Jonathan Sterne has termed a "speaker culture" (Sterne 2015: 113). The film's soundscape of technical communications and automated announcements, interwoven through montage, raises the question of whether the characters actually interact with human interlocutors or merely with automatically triggered answers stored on tape. Its references to telephone technology are inscribed further in its very scene design, with a Pacific Bell circuit switch room serving as a filming location, according to the IMDB trivia section:

> The seemingly endless Control Room where the android police try to corner THX and SRT, who find out LUH has been consumed for organ reclamation, was the circuit switch room of the San Francisco location of the Pacific Bell Telephone Company. Pacific Bell allowed George Lucas to shoot the film there, because the entire room and the hardware found there were about to be dismantled, as the phone company was switching to touchtone phone technology (IMDB 2019).

Lucas even named the title of the film after his San Francisco telephone number, 849-1138, where the letters THX correspond to letters found on the buttons for the digits 8, 4, and 9. Moreover, many of the electro-acoustic sound effects that populate the soundscape of the film are distilled from telephone dial tones, which editor and sound editor Walter Murch manipulated by applying compositional methods derived from musique concrete. The depiction of automatic speech systems as inhumane and anonymous is achieved largely by recreating or mimicking the user experience of early IVR systems: the messages and public announcements that are automatically triggered throughout the movie are repetitive and monotonous and leave no room for doubt that the citizens of the future society have to adjust to the system and not the other way around. Rewatching the movie almost half a century after its initial release, one cannot help but associate it with current AI-based public surveillance systems, such as China's Social Credit System.

3. "Speech is an Untapped Goldmine": The Adoption of AI in the Contact Center and Virtual Voice Assistants

Despite their still apparent limitations, recent speech recognition and synthesis systems, such as those used for voice assistants, sound more familiar and less robotic and anonymous than the mantra-like reminders and announcements that

populate the soundtrack in *THX 1138*. Early examples of automatic speech recognition (ASR) include pattern-based models for detecting a limited ensemble of spoken sounds such as digits and words, where the recognition of an uttered digit or word is determined by its correlation with a set of stored reference patterns (Davis et al. 1952: 194). Among the well-known early examples of such applications, Bell Laboratories' "Audrey" (Pieraccini 2012: 55-59) and IBM's "Shoebox" (Dersch 1962) were able to recognize spoken digits and, in the case of Shoebox, even a limited number of commands if spoken by a familiar voice.[7] "HARPY," a speech recognizer developed in the mid 1970s at Carnegie Mellon University as part of the first ARPA project on speech understanding research, was already able to recognize a vocabulary of 1,011 words (Lowerre 1976). In the late 1970s, IBM's Dragon system heralded a new era of ASR systems based on hidden Markov models, the descendants of which were used in most IVR systems from the 1990s onward (Pieraccini 2012).

As noted in the previous section, speech recognition research yielded the potential use of the human voice to control automated systems and to transmit information to them. Spoken language thus represented an alternative type of productive sound alongside DTMF tones in automated telephone systems. Moreover, the integration of voice control into major computer operating systems such as Windows or MacOS, not least in order to increase accessibility for visually impaired users, points toward the conversational systems that we now see used in current applications and platforms for smartphones and smart home devices. Today, the combination of automatic speech recognition, understanding and synthesis—now largely based on artificial intelligence approaches—is referred to as natural language processing. A crucial step toward this stage of extended voice agent interaction has been the application of machine learning and deep learning techniques, which mostly rely on learning algorithms based on deep neural networks (DNN). Compared with previous methods, following from the historical precursors in speech recognition and synthesis described above, DNNs allow for the analysis and processing of voice audio with a much higher level of accuracy and naturalness (cf. Mary 2018: 50). The improvements are primarily due to the general increase in processing power, the use of cloud computing, and the access to vast amounts of training data. This is also the reason why big tech companies have in recent years increasingly developed natural language processing and offered AI solutions for call centers and voice assistants for smartphone or home use.[8] Special apps and platforms, such as Amazon's Lex, Google's Dialog-

7 The name "Audrey" is a loose acronym of "automatic digit recognition."

8 These systems are increasingly based on a centralized internet infrastructure dominated by cloud-based services provided by a few major market leaders, Amazon (AWS), Google (Google Cloud), and Microsoft (Azure). The speech recognition models, the emotion analysis metrics, and

flow, Facebook's Wit.ai, IBM's Watson, and Microsoft's LUIS, offer considerably straight-forward solutions for creating conversational bots. Not surprisingly, one of the major professional domains of AI application is the contact center industry. Google, for instance, boasts that its cloud services provide "AI-powered virtual agents for the contact center, including phone-based conversational agents known as interactive voice response (IVR)" (Google 2019).

Call centers offer ideal conditions for the introduction of voice-centered AI technologies because they constitute, as shown in section 2, highly compartmentalized, process-oriented and automated conversational environments with a long history of human–machine integration. From the beginning, developers and vendors of IVRs have conceived systems to which both customers and agents must adapt. Now with the most recent examples of virtual assistants, we can observe this logic upheld and transformed into new semi-automated conversational settings: to ensure that the systems "understand" them, users need to adapt the way they talk and the words they use. Therefore, the processes of automating telephone work through IVRs and conversational AI can be better described in terms of what Hamid R. Ekbia and Bonnie A. Nardi (2017) have termed "heteromation," the "extraction of economic value from low-cost or free labor in computer-mediated networks"—in this case, labor performed by both call center agents and customers.

The contemporary contact center's core function does not fundamentally differ from its original, historical task of handling inquiries and improving customer satisfaction. However, the increase in online shopping and other forms of e-commerce has brought along a huge demand for virtual customer services, which coincides with the significant advancement in natural language processing capabilities and synthetic speech models over the past decade (Kopparapu 2015: 5). The use of these optimized systems promises the automation of not only call distribution and routing but also more individual customer interactions, such as complex three-factor account authentification, with the effect of further reducing the need for direct contact with human service agents—possibly until eventually conversations between humans will have shifted from the norm to the exception.[9] Since their emergence, IVR systems, which require long automated spoken menus for their extensive decision trees, have incurred criticism for being impersonal and annoying (cf. Smith 2016). A second reason for integrating intelligent personal as-

the design of synthetic voices that lie at the core of contemporary and future autonomous conversational agents are, thus, all dependent on the protocols and regulations determined by these corporates.

9 A parallel development has happened in the field of text-based chatbots, the performance of which is now convincing in standard use cases (Sheth et al. 2019), although as of now, most customer interactions still happen over the phone.

sistants is therefore to offer the customer the experience of a "personal," seemingly individual conversational behavior, with the goal of overcoming the perceived shortcomings of IVR systems, by hiding the underlying hierarchical structure from the perception of the customer. Apple's Siri, for example, has been branded from the start as a witty and fun-to-use application with personality to dissociate it from anonymous automated systems, such as IVRs.[10]

The use of automatic speech recognition to augment traditional IVR systems and replace human agents, however, is not the main purpose for introducing AI technology in the contact center. Rather, it is the tip of the AI iceberg that is generally visible or perceivable to the customer. In the contemporary contact center, we are very likely to find not only one but a growing number of different types of artificial intelligence solutions at work simultaneously. According to one of the industry's leading trade magazines, *Call Center Helper*, artificial intelligence solutions are used not only for the handling of calls but increasingly for the production of new insights about customers and call center agents by capturing data from customer interactions, applying big data analytics, predicting customer behavior or monitoring advisor performance (Call Center Helper 2018). An industry representative hence predicts that "our future with machines is going to be (and needs to be) one of partnership and enhancement, not sweeping replacement" (Call Center Helper 2019). Call centers usually have vast amounts of stored voice recordings at their disposal, which make them particularly suited for analytic AI applications, especially predictive analytics and speech analytics. As an industry white paper frames it, "Speech is an untapped goldmine." (CallMiner 2019: 5)

Predictive analytics allows call centers to generate valuable insights in real-time, such as a customer's willingness to pay off a debt, a customer service agent's effectiveness at addressing particular concerns, and a caller's overall sentiment and the actions likely to satisfy them given their history. Speech analytics, in turn,

> goes beyond recognition, interpreting not just the words a caller speaks but also the manner in which those words are spoken. [Also known as voice analytics, this technology] detects factors such as tone, sentiment, vocabulary, silent pauses, and even the caller's age, analyzing these factors to route callers to the ideal agent based on agents' success rates, specialized knowledge and strengths, as well as the customer's personality and other behavioral characteristics. (Stanley 2018, n.p.)

In particular, this concerns the backtracking of all available voice recordings for all sorts of analyses and the ambition to detect and analyze not only the semantic but also the emotional aspects of the human voice by exploiting methods of affective computing (Picard 1997; Jeon 2017). The bottom line of the current shift

10 "Siri" stands for "speech interpretation and recognition interface."

toward the integration of AI technology into the contact center is that it works only in part for the customer and primarily for the enterprise using it. What I want to argue is that the same goes for virtual voice assistants for the home, for which contact centers served as a testing ground for early adoption (Davis 2019). To this effect, the various uses of AI in the customer service industry point to a number of potentially invisible or concealed uses of AI for domestic voice assistants. Smart speakers with voice interfaces are branded as convenient interfaces to both local and cloud-based digital services. They are thus designed to simulate personality in order to be more fun to use. We should, however, not trick ourselves into thinking we are dealing with one and only one artificial intelligence alone—the workings of which are represented and condensed in the form of the artificial voice. Rather, we should realize that there are probably a dozen other AI systems listening in and analyzing the information of our voice data being transmitted to the providers' cloud servers. In the end, intelligent personal assistants work not merely *for* the users but *on* them. Domestic users and office workers embrace voice assistants for their convenience and efficiency in performing repetitive tasks such as web searches and daily routines. Businesses, tech corporations, surveillance states, and other actors, however, are competing to gain access to the users' voice itself, which is seen as a highly valuable data source—a "goldmine"—for AI-based analytics.

4. Conclusion

With the introduction of DTMF signaling in the 1960s, special-purpose telephone receivers were repurposed into general-purpose remote controls, resulting in a fundamental first step toward a long-ranging transformation of the telephone system from a mere medium of communication into a versatile medium of cooperation. Over the course of roughly two to three decades, Touch-Tone calling in conjunction with IVR systems slowly trained users in how to interact with remote automatic and semi-automatic information systems over the telephone network. Given that these technologies almost exclusively relied on all-acoustic interfaces and a small keypad, we can consider the mobilization of productive sounds to have ultimately paved the way for what could retroactively be called the first generation of everyday "online practices." Different iterations of productive sounds, as I have argued, in this way formed the basis of a slow transition from telecommunication to telecooperation: at first, in an essential and operational sense in the form of multifrequency signals; later, as voice work performed by call center agents, prerecorded messages, hold music and other design elements, which formed part of telephonic waiting loops and acoustic interfaces in automat-

ed interactive voice response systems; and finally, as conversational AI systems based on natural language processing.

As the introduction of Touch-Tone calling has shown, already the "old" media industries, especially the telecommunications sector, worked toward realizing a future based on networked information technologies and services. The automation of customer service calls revealed how infrastructural innovations laid the foundation for the emergence of new services based on both electronic and embodied "data practices" and how these transformations occurred in circuit-switched telephone networks before the growth of personal computers and the internet and well outside the computer industry. By tracing the relations between different technological agents and forms of labor within the cooperative assemblages of call centers, I have shown that the development of voice-related artificial intelligence systems should be seen as part of a larger history of human–machine interaction, the practices of which continue to shape the relations between users and contemporary voice assistants. This transformation occured not so much in the form of a disruptive revolution but in terms of historical continuities based on successive combinations and recombinations of (semi-)automatic man–machine systems and the sedulous infrastructuring, networking, and delegation of co-operative practices, ultimately leading to virtual call center agents and domestic voice assistants.

The use of voice assistants and smart speakers is reminiscent of the principles and practices of using a self-service call center system. Therefore, I tentatively like to frame them as call centers for the home. Moreover, in the coming years, voice control, especially in hands-free environments such as moving vehicles, is likely to become a ubiquitous and naturalized interface practice. In the contemporary contact center, managing and automating conversations to reduce labor cost and enhance efficiency is not the only motivation for embracing artificial intelligence solutions anymore; equally important is the analysis of user data for making predictions and producing new commercially exploitable insights. AI in virtual voice assistants is therefore used not only to create new ways of conveniently controlling our everyday tasks but also to data mine the control signals (i.e., the voice input) as exploitable customer data. Studying call center practices can therefore be a way to understand voice assistants, and their politics might thus best be explained by an uncanny pact of co-operation: One the one hand, voice assistants are devised to help us, and they do it well and will even get better as their skills improve. On the other hand, because virtual voice assistants transmit our digitized voice signals to remote cloud servers for processing, users are, metaphorically speaking, inviting into their homes and feeding nameless background AI routines with every conversation. The most common prerecorded pronouncement in call center systems is equally valid for virtual voice assistants: "Your call will be monitored."

Acknowledgements

This research has been funded by the German Research Foundation (DFG) as part of the A01 project of the Collaborative Research Center 1187 "Media of Cooperation" (Medien der Kooperation). I would like to express my gratitude to Kyle Stine for critical remarks and suggestions. I would like to express my gratitude to Kyle Stine for critical remarks and suggestions and Thomas Bjørnsten for valuable input to the paper. I would also like to thank Sheldon H. Hochheiser and Melissa Wasson of the AT&T Archives and History Center (Warren, NJ) for their generous support.

Bibliography

Aharon, Dan/Laqab, Daryush (2018): "Transforming the Contact Center with AI." Google Cloud Blog. https://cloud.google.com/blog/products/gcp/transforming-the-contact-center-with-ai/ (June 10, 2019).

Austin, John Langshaw (1975): How to Do Things with Words. Oxford: Oxford University Press.

Bodin, Madeline (2002): The Call Center Dictionary: The Complete Guide to Call Center and Customer Support Technology Solutions. Boca Raton: CRC Press.

Call Centre Helper (2018): "12 Top Uses of Artificial Intelligence in the Contact Centre." Call Centre Helper. https://www.callcentrehelper.com/12-top-uses-of-artificial-intelligence-in-the-contact-centre-123361.htm (June 10, 2019).

Call Center Helper (2019): "Artificial Intelligence in the Contact Centre: What You Should REALLY Know." Call Centre Helper. https://www.callcentrehelper.com/artificial-intelligence-contact-centre-should-know-142841.htm (June 10, 2019).

CallMiner Eureka (2019): "How AI Improves the Customer Experience. Real Use Cases of Engagement Analytics & Automation for Contact Center Success." CallMiner. https://learn.callminer.com/whitepapers/how-ai-improves-the-customer-experience (June 10, 2019).

Canalys (2019): "Canalys: Global Smart Speaker Installed Base to Top 200 Million by End of 2019." https://www.canalys.com/newsroom/canalys-global-smart-speaker-installed-base-to-top-200-million-by-end-of-2019 (June 10, 2019).

Dahlbom, C. A./Horton, Jr., A. W./Moody, D. L. (1949): "Applications of Multifrequency Pulsing in Switching." In: Trans. AIEE 68, pp. 392-96.

Davis, Jessica (2019): "Voice Assistants Coming to the Enterprise." InformationWeek. https://www.informationweek.com/strategic-cio/it-strategy/voice-assistants-coming-to-the-enterprise/d/d-id/1333642 (June 10, 2019).

Davis, K. H./Biddulph, R./Balashek, S. (1952): "Automatic Recognition of Spoken Digits." In: Journal of the Acoustical Society of America 24/6, pp. 637-642.

Dersch, W. C. (1962): "Shoebox: A Voice Responsive Machine." In: Datamation 8/6, pp. 47-50.

Ekbia, Hamid R./Nardi, Bonnie A. (2017): Heteromation, and Other Stories of Computing and Capitalism. Cambridge, Mass.: MIT Press.

Gillespie, Tarleton (2010): "The Politics of 'Platforms.'" In: New Media & Society 12/3, pp. 347-364.

Gopinath, Sumanth (2013): The Ringtone Dialectic: Economy and Cultural Form. Cambridge, Mass.: The MIT Press.

Green, Venus (1995): "Goodbye Central: Automation and the Decline of 'Personal Service' in the Bell System, 1878-1921." In: Technology and Culture 36/4, pp. 912-949.

Haigh, Thomas/Russell, Andrew L./Dutton, William H. (2015): "Histories of the Internet: Introducing a Special Issue of Information & Culture." In: Information & Culture 50/2, pp. 143-159.

Harr, J. A., E. S. Hoover/Smith, R. B. (1964): "Organization of the No. 1 Ess Stored Program." In: Bell System Technical Journal 43/5, pp. 1923-1959.

Hopkins, Harris F. (1960): "Push Button 'Dialing.'" In: Bell Laboratories Record 38/3, pp. 82-87.

IMDb (2019) "THX 1138 (1971)—Trivia." IMDb.com. https://www.imdb.com/title/tt0066434/trivia (June 10, 2019).

Jeon, Myounghoon (ed.) (2017): Emotions and Affect in Human Factors and Human-Computer Interaction. London; San Diego, Cal.: Elsevier/Academic Press.

Kopparapu, Sunil Kumar (2015): Non-Linguistic Analysis of Call Center Conversations. New York: Springer.

Lipartito, Kenneth (2003): "Picturephone and the Information Age: The Social Meaning of Failure." In: Technology and Culture 44/1, pp. 50-81.

Lowerre, Bruce T. (1976): "The HARPY Speech Recognition System." PhD Dissertation. Carnegie-Mellon University.

Mary, Leena (2018): Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition. New York: Springer.

Meacham, L. A./Power, J. R./West, F. (1958): "Tone Ringing and Pushbutton Calling: Two Integrated Exploratory Developments." In: Bell System Technical Journal 37/2, pp. 339-360.

Nexidia Interaction Analytics (2017): "AI-Powered Analytics Transform the Enterprise." www.nexidia.com/media/2522/whitepaper-using-ai-powered-analytics-jan-2017.pdf (June 10, 2019).

Noweck, H. E. (1961): "The Versatility of Touch-Tone Calling." In: Bell Laboratories Record 39/9, pp. 312-316.

Perez, Sarah (2019): "Over a Quarter of US Adults Now Own a Smart Speaker, Typically an Amazon Echo." TechCrunch. http://social.techcrunch.com/2019/03/08/over-a-quarter-of-u-s-adults-now-own-a-smart-speaker-typically-an-amazon-echo/ (June 13, 2019).

Picard, Rosalind W. (1997): Affective Computing. Cambridge, Mass.: The MIT Press.

Pieraccini, Roberto (2012): The Voice in the Machine: Building Computers That Understand Speech. Cambridge, Mass.: MIT Press.

Schüttpelz, Erhard (2017): "Infrastructural Media and Public Media." In: Media in Action 1/1, pp. 13-61.

Sheth, Amit/Yip, Hong Yung/Iyengar, Arun/Tepper, Paul (2019): "Cognitive Services and Intelligent Chatbots: Current Perspectives and Special Issue Introduction." In: IEEE Internet Computing 23/2, pp. 6-12.

Smith, Ernie (2016): "The History of the Call Center Explains How Customer Service Got So Annoying." Vice. https://www.vice.com/en_us/article/xyg4mn/the-history-of-the-call-center-explains-how-customer-service-got-so-annoying (June 10, 2019).

Soderberg, J. H. (1969): "Machines at Your Fingertips." In: Bell Laboratories Record A 47/7, pp. 199-203.

Stanley, Robert (2018): "A Comprehensive History of AI in the Call Center: From ACDs to Predictive Analytics and Beyond." CallMiner. https://callminer.com/blog/comprehensive-history-ai-call-center-acds-predictive-analytics-beyond/ (June 10, 2019).

Star, Susan Leigh/Bowker, Geoffrey C. (2002): "How to Infrastructure." In: Leah A. Lievrouw/Sonia Livingstone (eds.), Handbook of New Media: Social Shaping and Social Consequences of Icts, London: Sage, pp. 151-162.

Star, Susan Leigh/Griesemer, James R. (1989): "Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39." Social Studies of Science 19/3, pp. 387-420.

Sterne, Jonathan (2006): "The Mp3 as Cultural Artifact." In: New Media & Society 8/5, pp. 825-842.

Sterne, Jonathan (2012): MP3: The Meaning of a Format. Durham: Duke University Press.

Sterne, Jonathan (2015): "Space within Space: Artificial Reverb and the Detachable Echo." In: Grey Room 60, pp. 110-31.

Volmar, Axel (2017): "Formats as Media of Cooperation." In: Media in Action 1/2, pp. 9-28.