Siri, Give Me Back my Eye: From Audio Culture to Video and Back

Hassan Choubassi

The International University of Beirut
Michel Abi Chahla street, Mousytbeh
Beirut-Lebanon
hassan.choubassi@b-iu.edu.lb

Joe Elias

The International University of Beirut
Michel Abi Chahla street, Mousytbeh
Beirut-Lebanon
joe.elias@b-iu.edu.lb

Sahar Sharara

The International University of Beirut
Michel Abi Chahla street, Mousytbeh
Beirut-Lebanon
sahar.sharara@b-iu.edu.lb

With the innovation of the mass production media that resulted from the invention of the Gutenberg press more focus was given to visual media and listening was replaced by reading, “we were given an eye for an ear” according to Marshal McLuhan. With the invention of instantaneous media technology, virtual reality images imploded in the social in such a way that the visual images along with the audible sound produced a ubiquitous media with overwhelming presence. Yet again when this media connectivity became mobile audio sensory and voice commands added a layer of verbalism to media, an oral culture that was lost since the invention of the printing press is now regaining ground. Mobility added to connectivity required the use of all communication paradigms available to be used all together concurrently.

Tele-phonetic and audio controls are becoming essential in our interaction with machines, from the “Speech Interpretation and Recognition Interface” Siri to Alexa to Ok Google and Advanced Driver Assistance ADAS. Intelligent Virtual Assistant IVA and Voice User Interfaces VUI are changing social behavioural patterns and thinking modes in so many ways. We find ourselves forced to abandon the linear sequential media we know and halt into the core of the fragmented spherical feed, a feed of violence providing a huge amount of auditory and visual information. It is an act of violence, a feed that violates the individual's privacy and private spaces through manipulation, monitoring and surveillance. Audio information and oral modes of interactivity are adding a new layer of complexity to the already saturated scene of media communication.

Tele-phonetic. Voice command. Audio culture. Media augmentation. Mobile connectivity. Voice user interface VUI. Intelligent virtual assistant IVA.

1. INTRODUCTION

When I ask Siri to give me back my eye, “she” replies: “That may be beyond my abilities at the moment” little does she know, or maybe she knows a lot, about Marshall McLuhan's saying that we were given an eye for an ear. In his book “The Gutenberg Galaxy” he argues that printing press and phonetic alphabet transformed the mode of perception from the audible to the visual. Listening was replaced by reading and we were given an eye for an ear: “Civilization gives the barbarian or tribal man an eye for an ear and is now at odds with the electronic world” (McLuhan, 2002, p. 26). But yet

again, he claims that “Electric technology” brought us back to audio based society, compared to one that is visual based which started with the invention of the phonetic alphabet, “(...) Our world shifts from a visual to an auditory orientation in its electric technology” (McLuhan, 2002, p. 26).

Today, with the new technologies of mobile connectivity tele-phonetic media are regaining ground. Audio controls and virtual assistants (intelligent virtual assistant IVA or intelligent personal assistant IPA) are becoming essential in our interaction with machines, from Siri to Alexa to Ok Google, Advanced Driver Assistance ADAS,

GPS navigation devices and smart call centres. The voice-user interfaces VUI enables the users to interact with a virtual assistant through voice command they can ask direct questions, control smart devices and media triggers, in addition to managing calendar and emails (write, read, send and receive mails).

Today's media augmentation blurs the lines of what really constitutes a "medium". It has in many ways altered the way we think or act, as we find ourselves forced to abandon the linear sequential media we know and halt into the core of the fragmented spherical feed, a feed of violence providing a huge amount of auditory and visual information. It is an act of bombardment, a feed that violates the individual's privacy and private space through manipulation, monitoring and surveillance. Audio information and oral modes of interactivity are adding a new layer of complexity to the already saturated scene of media communication.

2. INTERACTIVE DIGITAL PROSTHESES

Technological changes in the field of audio communication and voice recognition, in addition to advancement in Artificial Intelligence (AI) devices allowed for tremendous changes in communication modes on the socio-political level. In an essay on the effects of technology on social urbanisation and mobility "The third Interval, a Critical Transition in Rethinking Technologies", in 1993, Paul Virilio traced the shift from the revolution of modes of transportation after the industrial revolution to the revolution of electronic communication and virtual reality. Virilio noted the changes that the new communication technologies are having on the urbanization of space and time and on the human body itself, and how people are forever locked to use devices that will enable them to go through their quotidian:

"The urbanization of real time entails first of all the urbanization of "one's own body," which is plugged into various interfaces (computer keyboards, cathode screens, and soon gloves or cyberclothing), prostheses that turn the over-equipped, healthy (or "valid") individual into the virtual equivalent of the well-equipped invalid. If the revolution of modes of transportation of the last century had witnessed the emergence and progressive popularization of the dynamic automotive vehicle (train, motorcycle, car, airplane), the current electronic revolution is now, in its turn, blueprinting the plan for the innovation of the ultimate vehicle, the static audio-visual vehicle, in other words, the coming of a behavioral inertia of the receiver-sender, or the passage from this fabled "retinal suspension" on which the optical illusion of cinematic projection was based, to the "bodily suspension"

of the "plugged-in human being." This becomes the condition of possibility of a sudden mobilization of the illusion of the world, of an entire world, that is telepresent at every moment (Virilio, 1993, p. 2).

Virilio is claiming that people are becoming like handicapped without their prostheses, the digital devices and sensorial equipment that allowed them to communicate to the machine are becoming essential tools of survival. But Virilio was talking about the pre-smart phone technologies, he was analysing the effects of digital static devices that provides connectivity but requires stationary. The only way to be connected to the Internet prior to 3G connection, was the cabled network that forbids mobility:

"modes of instantaneous transmission prompt the inverse, that of a growing inertia. Television and, especially, teleaction, no longer require human mobility, but merely a local motility. Telemarketing, tele-employment, fax work, bit-net, and e-mail transmissions at home, in apartments, or in cabled high rises, these might be called cocooning: an urbanization of real time thus follows the urbanization of real space. The shift is ultimately felt in the very body of every city dweller, as a terminal citizen who will soon be equipped with interactive prostheses whose pathological model is that of the "motorized handicapped," equipped so that he or she can control the domestic environment without undergoing any physical displacement. We have before us the catastrophic figure of an individual who has lost, along with his or her natural mobility, any immediate means of intervening in the environment. The fate of the individual is handed over, for better or for worse, to the capacities of receivers, sensors, and other long-range detectors that turn the person into a being subjected to the machines with which, they say, he or she is "in dialogue" (Virilio, 1993, p. 5).

With smart phones mobility is now possible along with connectivity and smart devices are everywhere on the move, on each and every individual, in cars, in trains or on planes, we are always connected and accompanied by our personal assistant, an intelligent virtual assistant IVA or intelligent personal assistant IPA. It is a software that assists the end users in their daily activities via voice commands and voice recognition. Those virtual assistants are able to understand oral speech and respond through a synthesized voice. Those capabilities and usage of virtual assistants are becoming more and more powerful with the introduction of smart mobile devices their use is intensifying rapidly; people are relying on them with an emphasis on voice-user interfaces. Those technologies are now introduced and being used by almost all company interfaces like Apple with Siri, Amazon with Alexa and Google with ok Google on

personal computers, smartphones, tablets and smart speakers.

The Voice User Interface VUI is the technology used by all speech applications. It allows humans to interact with the machine by simply talking to it. Voice commands and oral communication with computers are now part of our daily routine. Our digital machines are now able to transform text into speech or speech into a written text all the same. The visual phonetic alphabet is now merged with the verbal human voice. VUI technology is becoming a default setting as users find them very practical and time saving with minimum of efforts as they do not require proximity, physical tactility or visual disturbance; they are becoming more and more reliable as the margin of error is, since, reduced to a minimum. Those interfaces are not only based on computer science genius but also on social sciences of semiotics, linguistics and psychology. They also rely on product designers and user-experience experts, as they require comprehensive understanding of the target audience and the end users psychographics. The VUI technology is becoming very accurate and efficient as it is in synchronization to satisfy the user's psychographic modes and in accordance to the required function or service it is to provide, it is easy to use and it delivers all the guidance as it promises. A good VUI service organizes phone calls and eliminate unnecessary iterations to allow elaborate mixed initiative dialogs as flexible interaction strategy between human and machine to contribute what it is best suited at the most appropriate time. Speech applications are now well constructed to perform specific automated process to provide virtual information augmenting actual reality with digital parameters instantaneously, continuously and interactively.

3. FROM LINEAR TO FRAGMENTED

VUI has been added to all transportation vehicles, to smart homes automation systems and home appliances, and to computer interfaces of all operating systems. They are the main technological paradigms that allow communication with the virtual assistant of smart phone and home speakers or other audio controlled devices. VUI permits the users to voice their requests and converse without using any other sensorial elements, they don't have to use their hands or even look at the screen to be able to interact with their devices, VUI are also capable of replying to multiple commands at the same time, with appropriate feedback, precisely imitating a real conversation.

When this type of technology is available for all users in an affordable and accessible manner, the users will give away all traditional means of

interactivity in favour of it, they will abandon the narrative type of media in favour of the fragmented type of information which is eventually causing a great paradigm shift in the sound-image production and consequently in the sound-image perception at the collective level; information production is not linear anymore and the perception is not a process of interpretation of structural syntagmatic meanings but rather a collection of fragmented paradigms that will eventually lead to signification without any attention required from the subjects. Signification is not anymore a Saussurean collection of structural set of tokens that are juxtaposed in a rational order that will eventually lead to meaning, the interpretation of any set of sound-images does not require a continuous mode of attention (Saussure et al., 1986). Information is scattered and abundant, diverse and hybrid, the perceivers of new media augmentation are set to absorb all sort of information in so many formats. The sound media and oral commands are adding a layer of interactivity to the already saturated visual interfaces of communication machines. The need to be connected at all time required the use of as much of the five senses as possible, and oral commands are becoming indispensable to maintain this connectivity when the other senses are being used.

Fragmented media is becoming the norm, and people are used to such kind of information when interacting with each other or with the machine itself, linear and narrative media require a lot of attention while fragmentation requires a minimum efforts while providing a huge quantity of information. In his book "Understanding Media: The Extensions of Man" Marshal McLuhan distinguished between two types of media perception in reference to the amount of attention vs. participation they require from the audience, he classified media into Hot and Cool. Hot media, or linear media in our context, requires the full attention of the perceiver but with less participation to engage with the message. It demands little interactivity since it is direct and one-way and the user simply receives the message without any active participation. Radio, films, books, and visual images are considered hot media because they require engagement from the perceivers with certain senses (sight and hearing) without any relation to the content of the media. On the other hand, Cool media, or fragmented in our context, requires less attention and less engagement but would necessitate more interactivity and participation from the perceiver. Perceiving fragmented media demands less attention but more speed in shuffling and jumping from one topic to another between different media formats and the person will gain a different type of understanding of the meaning based on the variety of interpretations

derived from a Frankensteinish collage of an abundance of images.

“Hot media are, therefore, low in participation, and cool media are high in participation or completion by the audience. Naturally, therefore, a hot medium like radio has very different effects on the user from a cool medium like the telephone. (McLuhan, 1994, p. 25)

When it comes to sound and voice commands the Saussurean semiological system will deal with them as structural construct, they will be read, as any other text is being decoded and deciphered, they are signifiers in a system of signs. When those signs are fragmented in a spherical setting of matrixes the interpretation is not anymore based on a linear narrative, and the structural mode of image analysis will cease to function. The signifier is not anymore a representation of a signified but rather a continuous creation of a signified; the signifier is now the image of reality and at the same time reality itself.

The traditional duality established between the oral and the visual fields reduces the importance of sounds and the role of hearing in the semiological practice, and rules out all possibilities of inter-semiotic interpretations that are characteristic of the oral culture and all phenomena in which language, visual images and verbal sounds combine to provide codified meaning in such a way that the process of hearing, or perceiving sounds, is interpreted with an infinity of meanings, and signification becomes a “metaphysical trap” (Barthes, 1972). It is a decline to a lower dimension of a strict regulation of sound volume and pitch variations; representation and linearization into a signifying sequence of organized coded sounds. The final result is a semiological signification of the audible field as a negative gap between sounds and signs. The illustration of the audible field has always indicated that the human hearing adapts itself to any format, medium or language around a specific spectrum of sound-images such as human or humanoid voices or musical patterns. We are to rethink the relationship between sound and vision, between audio and video elements present in every set of communication paradigms through a set of semiological hitches that are contiguous to the audible field. The sphere of sounds and hearing in association with the codification of meaning in a language or semantics processes becomes ambiguous in the interpretation of a signification that differentiate between seeing things and speaking about them.

On another level of surveillance there also exist a paradigm shift in the manner of conducting monitoring and control, the metaphor of the Panopticon that Michel Foucault (Foucault, 1995)

talked about when criticizing modern capitalism is taking yet a new dimension of hegemony where new technologies of media augmentation allow for a deeper level of surveillance that is spherical and fragmented without any requirement of direct visual contact (Manovich, 2007). And when it comes to audio commands and oral communication with the machine, Virtual voice assistants pose a big concern on privacy matters associated with the microphone, as such feature implies that the device is always listening.

4. SOUND IMAGE AS REALITY

The physical status of the perceiver is a most important variable that will dictate the interpretation of the image, whether moving or static whether fast or slow whether paying attention or not. If we continue the same logic of analysis as Walter Benjamin in his essay “The work of art in the age of mechanical reproduction” (Benjamin and Arendt, 1999), we can say that with the industrial technologies of “mechanical reproduction” of information the perceiver will value the image in reference to its ubiquitous state, coming from its exhibition value and mass reproducibility. With the information technologies of digital reproduction the perceiver will value the image in reference to its instantaneous and interactive statutes coming from live broadcast and online virtual reality with a negation of the elements of physical space and present time and thus a creation of a sound-image that becomes itself reality. Furthermore, with mobile technologies of smart phones the perceiver will value the image in reference to its ubiquity, its instantaneity and its mobility altogether as a result of a mixture of criteria of the previous technologies coming from space augmentation and mobile connectivity. The three types of sound-images produced with three different types of technologies implies three different types of interpretation, in the first the sound-image is perceived as chronicle and (linear signification), in the second the sound-image is perceived as a fragmented and (spherical signification), in the third the sound-image is perceived as chronicle and fragmented together, a sound-image that is so present, ubiquitous, actual and virtual at the same time, an inflated image to the extent of violence, an aggressive image that negates the cleavage between the two realities and combines them together in an explosive manner.

Technology was always considered a drive of history and of social change (Smith and Marx, 1994); the technological tools of image production and media dictated the way the image is produced and for whom it is produced and thus how it is perceived. In his “Deep time of the media: toward an archaeology of hearing and seeing” (Zielinski and Druckrey, 2008) Siegfried Zielinski traced the

implication of technologies on the image production through its fractures; he conducts an archaeological quest into the latent side of media history that was neglected by so many scholars, and continues to build up a holistic view of the media using scattered and mostly forgotten technological inventions that shaped today's contemporary media. In modern history, the industrial technologies of media production instigated a great shift in image perception: the mass production of information, of culture and the arts led to the ubiquity of the image. When Karl Marx linked the machine of mass production to human social history, he suggested that the changes in the modes of production imply changes in social relations; as a response to Pierre-Joseph Proudhon's "The philosophy of poverty" (Proudhon, 2011) he wrote in "The poverty of philosophy" that:

"Social relations are closely bound up with productive forces. In acquiring new productive forces men change their mode of production; and in changing their mode of production, in changing the way of earning their living, they change all their social relations. The hand-mill gives you society with the feudal lord; the steam-mill society with the industrial capitalist" (Marx, 1992).

The continuation of this logic might as well be the "jet-mill" gives you society with digital globalization, or, not to sound ridicule or ridiculing Marx's idea, this means that the new technologies of digital media yield to yet another change in social relations with a global magnitude, society with global capitalism. The mass production of instantaneous digital information generates this hot-air balloon of media and images, of economy and money transactions of virtual spaces and avatars. For Zielinski "The media are now redundant" (Zielinski, 2013); In his " . . . After the Media" he analyses how the tools of digital communication imply a systemic mode of production that affects all kinds of image manifestation in art practice and theory, in culture and politics. "Media-explicit thinking is contrasted with media-implicit thought", for him there should be a distinction between online existence and offline being, on the back cover of that book he simply wrote one sentence "now that it is possible to create a state with media, media are no longer any good for revolution.... They have taken on a systemic character".

To use the same technological reference in a comparison between three distinct successive eras of image evolution, the engine would be the perfect example as it is the most influential technology of the modern times; the invention of the steam engine, and later on the four-strokes engine, and thus the ability to transform heat into movement, marked the beginning of mechanical mass

production and repetition. Then came the jet engine that enabled the possibility of putting a satellite in orbit and thus the beginning of instantaneous communication and mass production of information. And lately the combination of the two engines together in an explosive hybrid formula that kept the link to the virtual while in the actual, the smart phones enabled the free mobility in the actual keeping the feet on the ground while the heads are up in the skies in the virtual space of fantasies.

Actual reality is back again into the scene after a long period of incubation in the closed virtual realm of the static computer. The image of actual reality is an image of representation; a signifier of a signified, reality has always been the source or the origin of all images until the advent of the instantaneous mode of production with digital media that negated the signified. The instantaneous image of the digital era is a simulation (Baudrillard, 1995), a creation of a new reality inside the virtual, a signifier without a signified, actual reality in that sense was lost and neglected; actual reality as the natural source of all representation. For Theodor Adorno nature is the origin of all art representations. In his "Aesthetic Theory, Theory and History of Literature" he wrote on natural beauty:

"The pure expression of artworks freed from every thing-like interference, even everything so-called natural converges with nature just as in Webern's most authentic works the pure tone, to which they are reduced by the strength of subjective sensibility, reverses dialectically into a natural sound: that of an eloquent nature, certainly its language, not the portrayal of a part of nature. The total subjective elaboration of art as a nonconceptual language is the only figure, at the contemporary stage of rationality, in which something like the divine language of creation is reflected, qualified by the paradox that what is reflected is also blocked. [...] If the language of nature is mute; art seeks to make this muteness eloquent" (Adorno, 1997, p. 101).

Many philosophers in relation to other types of images treated sound, noise and music differently. For Schopenhauer music has different standing among all art forms or any other man-made products, music for him is not a signifier that represents a signified, not an image of reality in that sense but it is reality itself, uniquely revealing the essence of the "in itself" of the world. Unlike all other images or art products that signify or mimic the Ideas, music is an expression of the "will" itself, qua-thing in itself, bypassing the prerequisite of the "Ideas" completely (Schopenhauer, 1966a, p. 285). On the other hand, noise, is the most disgraceful signifier of all: "noise is the most impertinent of all forms of interruptions. It is not only an interruption, but also a disruption of thought" (Schopenhauer,

1927). The perception of noise or the endurance to noise is a sign for the intellectual level of people:

“I have for a long time been of opinion that the quantity of noise anyone can comfortably endure is in inverse proportion to his mental powers, and may therefore be regarded as a rough estimate of them” (Schopenhauer, 1966b, p. 30).

For Schopenhauer the degree of perception of the arts is a unit of measurement for the quality of those arts, as he talks about “pure perception”. Nevertheless for him they are inferior to music - a direct manifestation of “will”. Music to Schopenhauer is the highest form of art (Schopenhauer, 1966a).

What would Schopenhauer think of voice command and VUI? It is, for sure, not music. If we consider voice command technology a mere representational product of mass produced information of the augmentation era things will have to follow the same logic of analysis of media fragmentation that requires less attention but more participation and interactivity. It is not noise either, voice command and all tele-phonetics will never reach the level of “pure perception”. Sound of this sort falls in the same category of obscenity, as per Jean Baudrillard's idea of the excess of information, (Baudrillard and Petit, 1998) where images become a pornographic simulation of reality.

5. CONCLUSION

Tele-phonetic, Voice User Interfaces, intelligent virtual assistant and voice controls are now part of our daily life and central to our interactivity systems that we have with the machine. Siri and Alexa are our friends, if not lovers, that we cannot live without. Advanced Driver Assistance ADAS on GPS maps in our cars help us get through our day in an ever-demanding urban environment, our social interaction is changing accordingly, our behavioural patterns and thinking modes are now directly related to those technologies that we are taking for granted. We are compelled to withdraw from our conventional modes of perception in favour of the new fragmented ones. The linear sequential media is seldom being in use and we are now driven into the core of the fragmented non-narrative feed, a feed of violence providing a gigantic amount of information, whether visual or auditory. It is a feed that violates the individual's privacy and private spaces through manipulation, monitoring and surveillance. Audio information and oral modes of interactivity are adding a new layer of complexity to the already saturated scene of media communication.

Media augmentation and oral voice commands thrust us into a reality of uncertainty where media dictates what we are and what we need. It has in many ways altered the way we think or act; we are bombarded by a massive amount of information that overloads our senses: seeing, hearing, talking and touching. The level of surveillance that these media enforce is also augmenting in quantity and quality, the fact that we now have a microphone that is always open and ready to take command raises a lot of questions on what are those devices listening to, to what extent the audio information collected by those capturing devices are being used in advertising and marketing schemes. Many myths of audio controls and monitoring are circulating around social media, people feel like they are being watched and heard at all times, when they receive advertisements related to something they just talked about an hour ago makes them paranoid and suspicious. The Panopticon is alive and heavily armed with all sorts of media tools of sensors, and capturing devices.

6. REFERENCES

- Adorno, T.W., 1997. Aesthetic theory, Theory and history of literature. University of Minnesota Press, Minneapolis, Minn.
- Barthes, R., 1972. Mythologies. Farrar, Straus and Giroux.
- Baudrillard, J., 1995. Simulacra and Simulation, 1st Edition. ed. University of Michigan Press.
- Baudrillard, J., Petit, P., 1998. Paroxysm: interviews with Philippe Petit. Verso, London; New York.
- Benjamin, W., Arendt, H., 1999. Illuminations. Pimlico, London.
- Foucault, M., 1995. Discipline & Punish: The Birth of the Prison, 2nd edition. ed. Vintage Books, New York.
- Manovich, L., 2007. The poetics of augmented space, in: Space Time Play. Birkhauser, Berlin.
- Marx, K., 1992. The Poverty of Philosophy. Intl Pub, New York.
- McLuhan, M., 2002. The Gutenberg galaxy: the making of typographic man, Reprinted. ed. Univ. of Toronto Pr, Toronto.
- McLuhan, M., 1994. Understanding Media: The Extensions of Man, Reprint edition. ed. The MIT Press.
- Proudhon, P.-J., 2011. The Philosophy of Poverty. Theophania Publishing.
- Saussure, F. de, Bally, C., Sechehaye, A., Riedlinger, A., 1986. Course in general linguistics. Open Court, LaSalle, Ill.
- Schopenhauer, A., 1966a. The world as will and representation, Vol. I. Dover Publications, New York.

- Schopenhauer, A., 1966b. *The World as Will and Representation*, Vol. II, 2 edition. ed. Dover Publications.
- Schopenhauer, A., 1927. *Studies in Pessimism*. Modern Library.
- Smith, M.R., Marx, L., 1994. *Does Technology Drive History? The Dilemma of Technological Determinism*. The MIT Press, Cambridge, Mass.
- Virilio, P., 1993. *The third Interval*, in: *Rethinking Technologies*. University of Minnesota Press, Minneapolis.
- Zielinski, S., 2013. *[...After the Media]: News from the Slow-Fading Twentieth Century*, Reprint edition. ed. Univocal Publishing, Minneapolis, MN.
- Zielinski, S., Druckrey, T., 2008. *Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means*. The MIT Press, Cambridge, Mass.