Dialogue with a Monologue: Voice Chips and the Products of Abstract Speech

Contact
Natalie Jermijenko
Center for Advanced Technology
New York University
710 Broadway, 12th Floor
New York, New York 10003 USA
+1.212.998.3382
+1.212.995.4122 fax
nat@cat.nyu.edu

This paper argues that voice chips and speech recognition chips can be used as a unique analytic tool for understanding the complex techno-social interactions that define, imagine, and produce new products. Using these chips as an in situ instrument allows a focus on products in their actual context of use, capturing the multiple interpretations of new technologies, and a method to analyze their failures and successes in human machine interaction. It is the use of voice that is direct evidence of the interactive, particularized and social aspects of products that are traditionally underrepresented in the attempts to understand technological innovation, design, and deployment.

The first part of the paper examines the use of integrated circuits that produce speech in consumer products, commonly called voice chips. The goal is to document what these products actually say and to try to understand what the voices of these products represent, specifically, what they say about techno-social relations. The paper describes how voice chip technology differs from other talking hardware of the recording and communications industries, and places it in a unique social and functional position: and provides insights into the possibilities of ubiquitous computational devices more generally. This section includes a survey of the voice chip patent literature; samples the products currently on the market; and investigates how the voices of these products can be interpreted as speech and interaction, drawing largely upon Suchman's examination of human-machine interaction. I conclude this section by using the chips voice to question their performance of abstract speech, and preprogrammable interaction, and therefore what actually happens in the realworld context when we attribute speech and agency to technological products.

The second part of this paper introduces a preliminary examination of the opposite techno-social phenomena: what we say to our things (rather than what they say to us). Using speech recognition chip sets, which enable relatively widespread and cheap speech recognition to be embedded in devices as a secondary function (e.g. cell phones), we can hear and examine what we say to our devices. Taken seriously as speech acts we can recognize the social position our address conveys. In other words: now that we can speak to our things, what do we say? And, furthermore, what do we mean? Because there are not many of speech recognition engines deployed in distributed products currently, the method we have used to survey a range of applications is by hosting a competition in which entrants proposed speech recognition interfaces to existing product. Just under three hundred designs were submitted and are available on the Web site www.cat.nyu/neologue. These proposed applications are analyzed in terms of the technological desires, expectations, and hopes they embody: particularly popular is the desire for social and individual envisioning and regulation; and there clearly stated within these proposed product interfaces explicit desired social transformations. This initial analysis is presented in order to set up some the preliminary ideas and interpretation, so that as the speech recognition chips become more widely distributed we can tune in to this particular historical moment and hear what it is we expect, want and bring to our human machine interaction. Listening to our daily interactions with products can work to contest and complicate the dominant methods used to describe technological trends and patterns of product innovation: demographically driven and massified market research and the capture of consumptive behaviors at point of purchase.

The database of voice chip and speech recognition products, patents, and sound files is available at www.cat.nyu/neologue, including instructions on how to contribute to material, from products to further analysis.

Voice: a social technology

Voice is the icon of person. It is the icon of the political agent. "To be given a voice" is how we understand the fundamental unit of democracy, voting or being represented. It is the recognition of each person and also the device for interpolating a subject into society.[1] In short it is the fundamental device of sociality and therefore interaction. Used in contrast to techniques and technologies, the voice is a responsive and ephemeral social device. The predetermined functions of products, manufacturing systems, word-processing software and other work-related technologies symbolize the stable, predictable, and material aspects of society, while the voice is reserved as the device that is claimed to define human-ness, expressing emotion, negotiating, conversing, and ultimately, having agency. In fact, the preconditions for individuation and socialization rely if not directly on the voice, then at least symbolically. Individual agency and free will are both preempted by the voice and operationalized through the voice. All notions of the social are somehow tangled with the voice.

Further, a voice is always understood from a social position.[2] Thus, if talking is the act of sociality then the product must talk from its social position. Or conversely perhaps the products words are understood from its place in the social world. Giving a voice, gives a political presence – to be counted, understood, or at least listened to. Because voice is situated and local (the same words can mean different things in different contexts), voice chip products articulate the tension between the product as a mass market phenomena and its actual incarnation into an individuals daily activity and meaning making.

Hearing Voices

Given this theoretical context we ask: What does technology have to say for itself? When hardware has a voice, what does it say?

Talking hardware has existed since before the time of Thomas Edison, who is generally credited with a having invented the phonograph around 1877, when Alexander Graham Bell's telephone learnt to talk. The proliferation of talking hardware since has bought the recording industry, the broadcast industry, and the multimedia industry. Our exposure to voices (and other com-

municative sounds) that emanate from inanimate objects has become a significant part of our daily interactions: from radios to talking elevators, answering machine messages, and prerecorded music, television, automated phone menus, automatic teller machines, alarms and alerts, each of which, we will show, speaks in a language or dialect that makes little distinction between music, sound effects, and articulated words.

There are, however, interesting distinctions to make between the voice chips, the concern of this paper, and noisy hardware more generally. Voice chips refers colloquially to: Texas Instrument TSP50C04/06 and TSP50C13/14/19 synthesizers; Motorola MC34018 or any other "speech synthesis chip implemented in C-MOS to reproduce various kinds of voices, and includes a digital/analog (D/A) converter, an ADPCM synthesizer, an ADPCM ROM that can be configured by the manufacturer to produce sound patterns simulating certain words, music or other effects.[3]

The voice chip differs from other technologies of automated sound production in that it technically offers autonomous voices, as opposed to broadcast voices, that is, voices which are not necessarily associated with a performer or any other pre-established identity. These chips present local talk in independent products that need not make a claim to belong to an identity, or to the faithful reproduction of someone elses voice. In fact their sound quality has effectively precludes this. However, the I in "I'm sorry, I could not process your request" or the "I will transfer you now" voice of the automated operator claims agency by using the first person pronoun.[4] Presumably, the machine is referring to itself when saying I, because it is not identifiably anyone else. [5]

Attributing agency to technologies is a theoretical strategy that has been used by others to better understand the social role of technologies.[6] It is a strategy that dislodges the immediate polarization of techniques and society, a strategy that refuses reduction to a situation that is merely social or only technological. Latour bases his Actor Network Theory, a theory that regards things as well as people as actors in any socio-technological assemblage, on the ability of humans and non-humans to swap properties. He claims that every activity implies a generalized principal of symmetry or, at least, offers an ambiguous mythology that disputes the unique position of humans. Callon and Law have also explored non-humans as agents, but their strategy starts with an indisputable agent (a white male scientist) and strips away his enabling network of humans and non-humans to demonstrate that his agency, his ability to act as a white male scientist, is distributed throughout his network of people, places, and instruments.[7] Even a more traditional theory like technological determinism rests on the assumption that technology has an agency apart from the people who design, implement or operate it, and hence can determine social outcomes. Voice chip products take these ideas literally and actually attribute, with little academic debate or contest, the defining human quality of speech to technology. Voice chips have humbly preempted the theory.[8]

The voices of chips differ from those of loudspeakers, TV/radio, and other broadcasting technologies in the social spaces they inhabit. Although radio and TV have become so portable that their voices can emanate from any vehicle, serving counter or room, voice chip voices, by virtue of their peripheral relationship to the product, inhabit even more radically diverse social spaces. The identity of the voice that emanates from TV and radio reminds us that it is coming from elsewhere "for CBS News," "It is 8 oclock GMT; this is London." And although Channel 9 is not a physical place, its resources and speech are organized around creating its identity, as an identifiable place on the dial. The voice chip that tells you "your keys are in the ignition" is not creating a Channel 9 identity, however. Its identity is "up for grabs," not quite settled, it speaks from a position of a product in the social space of daily use.

Similarly, recording media and hardware refer to what they record. We know we are listening to someone when we listen to an Abba CD. And although it is the tape-recorder in the car that produces the sound, we claim to be listening to the Violin Concerto itself. The tape recorder as a product does not itself have a voice, it never pretends to sing, speak, or synthesize violin sounds itself. The recording industry and associated technologies, born at a very different historical moment from voice chips, came out of the performance tradition.[9] Its claim to represent someone, from the earliest promotions using opera singers, to contemporary mega stars, has focused the technologies around "fidelity" issues. Additionally, telephones, telephonic systems, and the telecommunications industry, motivated by communication imperative, prioritize real-time voices passing to real-time ears, over fidelity. Simply stated, it is an industry that puts technologies between people, things to communicate through, "overcoming the tyranny of distance."[10] Invisible distance and seamless technology, reflect the recording industrys ambition to "overcome the tyranny of time," enabling people to duplicate the performance regardless of when or where it was originally performed. Voice chips and their inferior sound quality, do not refer beyond themselves. Their position in a product becomes their position as a product.

The Distribution of Voice Chips

Voice chips provide the opportunity to add "voice functionality" to the whole consumer based electronics industry. They are the integrated circuits that can record, play, and store sounds, and more importantly voice. They are the patented chips that play "Jingle Bells" in the Hallmark greeting card.[11] They are the voice in the car that reminds you "Your lights are on." [12] They are the technology that makes dolls that say "Meet me at the Mall,"[13] and give products ranging from picture frames to pens.[14] The well

sung virtues of integrated circuits (chips) is that they are cheap, tiny, and require little power. Smaller than a baby's fingernail, they have the force of a global industry of behind them and an entire economic sector invested in expanding their application. Technically, they can be incorporated into any product without significant changes in their housing, their circuit design, power supply, or price. Wherever there is a flashing light, there could instead, or as well, be a voice chip.

Although most computers can record and play voice, the voice chip is different in that it is dedicated solely to that function. The same integrated circuit technology of calculators and computers allows this tiny package to be placed ad hoc, in consumer devices. Their development exploited the silicon chip manufacturing processes and its dedication to miniaturization. With sound storage capacities ranging from seconds of on board memory to minutes and hours of recording time when configured with memory chips, they were conceived to enable voices in existing hardware, to be incorporated into products. They are the saccharin additive of consumer electronics.[15] They were first mass marketed in 1978 by Texas Instruments though they had existed in several forms before that. It was not until seven years later, in 1985, that the Special Interest Group in Computer-Human Interface (SIGCHI) of the American Computing Machine (ACM) professional society, mobilized an entire community to break off into their own conference from other more general computing conferences. This historic moment, which crystallized a discussion in design communities on the Human-Computer Interface as a site of scientific investigation, differs from earlier formulations of this interface, such as Engleberts human augmentation thesis or Turings standing-in-for ideal, but dominates still. This site, the liminal zone where people and machine purportedly interact is where the voice chips were intended to reside. The voice chips arrived to mediate, even to negotiate, this boundary. Voice chips promised to make hardware "user friendly," a phrase that defines the technical imagination of the time, by turning the person into an interchangeable standardized "user" and attributing a personality (i.e. friendliness) to the device. In this context the problem for designing user-friendly devices begins with the assumption that the hardware has agency in the interaction. Writes Turkle:

> Marginal objects, objects with no clear place, play important roles. On the lines between categories, they draw attention to how we have drawn the lines. Sometimes in doing so they incite us to reaffirm the lines, sometimes to call them into question, stimulating different distinctions.[16]

Marginal Voices

Finally, before listening to the voices themselves, I want to emphasize the peripheral relationship of the voice chip to the product. It is the position of the voice chip, as marginal, not particularly intended to be the primary function the product that increases the present curiosity in it. The motor vehicle, for example is not purchased primarily for its talking capacity, and pens that speak are useful for writing. This marginality gives voice chips a mobility to become distributed throughout the product landscape and mark, like fluorescent dye, a social geography of product voices.

The chips are usually deployed, to borrow from the economic sense of the term, for their marginal effects, to give one product (e.g. an alarm system) some marginal benefit over a competing product. However, the chips are not evenly distributed throughout competitive markets, (e.g. consumer electronics) in the manner one would expect for the propagation of a low-cost technical innovation driven by market structure alone. Although consumer preferences are often claimed to have a causal determination on the appearance or disappearance of marginal benefits, it is difficult to see how the well-developed paths of product distribution have the capacity to communicate those preferences developed after the point of purchase. Lending the market ultimate causality (or agency) ignores the specific experience of conversing with products, the micro interactions that enact the market phenomenon, and occludes the attribution of agency to the voice chip products, in so much as these products speak for themselves. The voice chip products themselves have something to say, although their voices are usually ignored. In this paper we will not examine voice chip products in the interactions of daily use, as contrapuntal to market descriptions, however by recognizing the social assumptions which determine their physical design, we frame the imagined interactions and social worlds in which these products make sense.

Finding the Voices

The marginality of the product makes it difficult to systematically study. Neither of the two largest manufacturers of voice chips of various types (Motorola and Texas Instruments) keep information on what products incorporate this technology, partly because they can be configured in many different ways, not necessarily as voice chips, and partly because products that talk are not a marketing category of general interest. This paper traces voice chips in two ways: firstly via the patent literature, and secondly through a more ad hoc method of searching catalogues, electronics, toy and department stores to compile a survey of products that were available at the time of my year long study (June 1996 to June 1997).[17]

What is initially observable from the list of products and patents that contain voice chips is that there is no systematic relationship between the products that include voice chips and the uses or purposes of those products. Except for children's toys, no one market sector is more saturated with talkative products than another. These chips are distributed throughout diverse products. However, we can view the voices as representatives, as in a democratic republic where voices are counted. Just as in a republic each citizen has a vote but most chose not to exercise it, likewise, most products could incorporate voice chips but most do not, so we will count what we can.

What do voice chips say?

A review of the patents literature yielded a loose category scheme, or a typology, not by where the voice chips appeared, but by what they said. The patents themselves hold a peculiar relationship to the products: For only two of the products on the market did I find the corresponding patents, the CPR device[18] and the recordable pen.[19] Though patents do not directly reflect the marketed products, they do represent a rather strange world

of product generation, a humidicrib for viable and unfeasible proto-products. Patents track how products have been imagined and while they do not by any means demonstrate market success, they do reflect a conviction of their worth, being invested in and protected. Patents are a step in the process of becoming owned, therefore worth money, and thereby demonstrate how voice, a social technology, becomes property.

There are as of March 2001 only 84 North American patents that include a voice chip. Of these 34 were issued in the year 1996-7, approximately 15 since, and the remainder in the previous five. In the context of the patent literature, the first thing to note is that this is a very small number, compared, that is, to the integrated circuit patent literature more generally. The question "why not more?" we will return to later. The federal trademark office offers a suggestive list of speech invoking names, including: who's voice; provoice; primovox; ume voice; first voice; topvoice; voice power; truvoice; voiceplus; voicejoy; activevoice; vocalizer; speechpad; audiosignature. These nomickers provide another introduction into how the voice is conceptualized in the realms of intellectual property. However the voice chips themselves seem to fall into the following categories: (a) Translators, which range from reporting and alerting to alarming and threatening and include interactive instructional voices; (b) Transformers, which transform the voice; (c) Voice as Music, that makes speech indistinguishable from music or that present voice as sound effect; (d) Locating Voices, speaking from here to there about being here; (e) Expressive Voices, expressing love, regret, anger, and affection (f) Didactic voices and Imitative voices, mainly as in the educational and whimsical children's toys; (g) Dialogue Products, which explicitly intend to be in dialogue with the user as opposed to delivering instructions to a passive listener.

The product and patents often exist in more than one of these categories; for instance, the Automatic Teller Machine will not only apologize (expressive) for being out of order but will also simply function to translate the words on the screen into speech. This said, the categories remain, for the most part, distinguishable and useful.

Translators

A large category, this is the voice that translates the language of buzzes and beeps into sentences whether English, French or Chinese. A translator is a chip that translates the universal flashing LED, the lingua franca of the peizo electric squeal, the date code, the bar-code, the telephone ringer adapter that translates that familiar ring, the tingling insistent trill of an incoming call, into "a well known phrase of music"[20] an approach that has since become popular in cell phones which this function finds a use in differentiating who's phone is ringing, or the unrelated patent that translates the caller identification signal into a vocal announcement. Within the translators there are distinct attitudes, for instance, the impassive reporting, almost a voice of nature. This is exemplified by the patent for the menstrual cycle meter. The voice reports the date and time of ovulation, in addition to stating the gender more likely to be conceived at a particular date or time during a woman's fertility cycle. Another example is the patent for the train defect and enunciating system, that "reports detected faults in English." These chips speak with a "voice of reality," reporting "fact" by the authority of the instrument that triggers them.

The other types of translator are more urgent than reporting. They raise alarm and expect response. They are less factual, more contestable perhaps. Take the "Writing device with alarm,"[21] an "invention which relates to a writing device which can emit a warning sound-or appropriate verbal encouragement — in order to awaken a person who has fallen asleep while working or studying"; or the baby rail device which exclaims "the infant is on the rail, please raise the rail,"…and then if there is no subsequent response from an attendant caregiver raises it automatically.[22] A product on the market that will politely tell you if there is water on the ground is pictured in figure 2.

These voice chips ask for and directs the involvement of their humans counterparts, they assume interactive humans. These chips articulate not only simple commands but series of instructions as well. The CPR device[23] (see figure. 3), guides the listener through the resuscitation process. And finally, these chips translate menus of choices into questions. The car temperature monitor that asks the driver "Would you like to change the temperature?" translates from the visual menu of choices but in the process also takes over the initiating role. What is lost or gained in the translation generates many questions: Does translating from squeals to a more articulate alarm make it any more alarming; how do spoken instructions transform written instructions? We will try to address these questions later.

There is an notable set of aberrant but related patents that exist in this category: The Alarm system for sensing and for vocally warning a person approaching a protected object[24]; The Alarm system for sensing and for vocally warning of an unauthorized approach towards a protected object or zone[25]; and the Alarm system for sensing and for vocally warning a person to step back from a protected object.[26] What seems almost a turn of a phrase to get three separate patents, has little technical consequence: the second patent has the extra functionality to detect authorized persons (or their official badge), and the third can, but need not, imply a different sensor perhaps, but each implies a different attitude. Although all patents are contestable, patent attorneys typically advise that you would not successfully win as separate patents an alarm system that warned at 15 feet from one that

alerted at two feet. The novel use being patented here depends on the wording, the phrasing of the instruction that determines the arrangement of the sensor and alarm/voice chip. On the strength of a differently worded warning the importance of the technically defined product description seems to have diminished. Perhaps ElectroAccoustic Novelties, the owners of the patents, have a linguist generating an alarm system for other phrases. These patents seem to be articulating the semantics of the technology. The intentionality of the system is its voice.

The transformers
The transformers are distinct from the patents that translate the voice. They translate in the other direction, not from the buzzes and squeals to spoken phrases but from the human voice to a less particular voice. For instance: to assist the hearing impaired, the chip that transforms the voices into the frequency range which still functions, usually into a higher frequency; or the "Electronic Music Device" effecting a "favorable musical tone." The voice tone color can be imparted with a musical effect, such as vibrato, or tone transformed.[27]

Into this category fall childrens products like "YakBak," popular in the 1997-99 seasons which plays back a childs voice with a variety of distortions, and the silicon-based megaphones that allow children to imitate technological effects, or sound like machines. These are voice mask for putting on the accent of techno dialect. The socializing voices broadcast on radio, and TV, the voices of authority heard over public address systems, and the techno personalities of androids and robots are practiced and performed in playing with these devices. This is also category of voice chips that is concentrated in products for the hearing impaired or otherwise disabled, and for children. These transforming devices act as if to integrate these marginalized social roles into a socio-technical mainstream.

Speech as Music
Many of the patents that are granted specifically collapse any difference between music and speech. This contrasts with the careful attention given to the meaning of the words used in the alarm system family of the Translators. An explicit example is the business card receptacle, which solves the problem of having business cards stapled onto letters making them more difficult to read, and provides an "improved receptacle that actively draws attention to the receptacle and creates an interest in the recipient by use of audio signals, such as sounds, voice messages, speech, sound effects, musical melodies, tones or the like, to read and retain the enclosed object."[28] Another example is the "Einstein" quiz game that alternately stated "Correct, you're a genius!!" or sounded bells and whistles, when the player answered the question correctly. This interchangeability of speech and music is common in the patent literature presumably because there is no particular difference technically. In this way patents are designed to stake claims – the wider the claim the better. The lack of specificity, and deliberate vagueness in genre of intellectual property law contradicts the carefulness of copyright law, the dominant institution for owning words.

Local talk from a distance
One would expect chips that afford miniaturization and inclusion in many low power products to be designed to address their local audience, in contrast to booming public address systems or broadcast technologies. However, several of these voice chip voices re-circulate on the already established (human) voice highways, imagined to transmit information as you or I would. The oil spill detector[29] that transmits via radio the GPS position of the accident, or the cell phone based automatic emergency vehicle location system which reports the latitude and longitude into an automatically dialed cell phone.[30] These are examples of a voice chip standing in for, and exploiting the networks established for humans, transmitting as pretend humans. This class of products, local agents speaking to remote sites, are curious because the information can easily be transmitted efficiently as signals of other types. Why not just transmit the digital signal instead of translating it first into speech? The voice networks are more public access, more inclusive, if we count these products as part of our public, too. The counter example, of voice chips acting as the local agent to perform centrally generated commands, is also common, as in the credit card actuated telecommunication access network that includes a voice chip to interact locally with the customer while the actual processing is done at the main switchboard. Although the voice is generated locally, the decisions on what it will say (i.e. the interactions) are not.

Expressives
The realm of expressiveness, often used to demarcate the boundaries between humanity and technology, is transgressed by the voice chips. There are, of course, expressive voice chips ranging from: a key ring that offers a choice of expletives, swear words and curses; the portable parent that plays stereotypical advice and parental orders; the array of Hallmark cards that wish you a very happy birthday, or say I love you. These expressives applications also remind us of the complexities of interpreting talking cards. The meaning of these products is of course, dependent on the details of the situation rather than the actual words being uttered: who sent the card, when; or what traffic situation preceeded the triggering of the key ring expletive.

These novelty devices lead into the most populous voice chip category: those intended for children. The local toy department store, Toys R Us, currently has seven aisles of talking and sound making products, approximately 45 different talking books alone, in addition to various educational toys, dolls and figures that speak in character. The voices are intended for the entire age range from the earliest squeaking rattles for babies to strategy games for children 14 years of age and up. For example the Talking Battle Ship in which you can hear the Navy Commander announce the action as well as "exciting battle sounds." The categorization of the multitude of toys extends far beyond expressive types; from the encouraging voices inserted in educational toys: "Awesome!," "No, try again," or "You're rolling now" in the Phonics learning system, the "Prestige Space Scholar," and "Einstein's Trivia" game; the same recordable voice

chips, used for executive voice memo pads, are for children placed in pens, balls and "YakBaks" (walkie talkies for talking to yourself); then there is the multitude of imitative toys that emulate cute animals, non-functional power tools and many trademarked personae from Tigger and Pooh to Disneys recent animation characters, Sampson and Delilah, Arial the mermaid, and others.

This listing demonstrates a cultural phenomenon that enthusiastically embraces children interacting with machine voices, and articulates the specific didactic attitudes that are projected onto products. These technological socialization devices have already been subject to analysis, for instances Turkles study of children attitudes towards interactive products.[31] Barbie, for instance, was taken very seriously for what she had to say about the most polarized notions of gender she embodies. Since Barbie's introduction in 1957 she has been given a voice three times (each with slightly different technology), her most controversial voice during the 1980s was censored for saying "Math is hard." This controversy rests on the assumption that voice chips are social actors and do have determining power to effect attitudes, in this case a young Barbie players attitude to math.

Although Barbie is currently silent, a myriad of talking dolls remain, from Tamagachi virtual pets, with their simple tweets, to crying dolls that ask to be fed, and an ever increasing vocabulary of robotic dolls creatures. The utility patent literature continues to award new and novel applications in this area. One of the new voice chip patents is for a doll that squeals when you pull her hair (dolls that cry when they are wet or turned upside down are technically differentiated by their simple response triggers).[32] There is also a new doll patent that covers electronic speech control apparatus and methods and more particularly for... talking in a conversational manner on different subjects, deriving simulated emotions...methods for operating the same and applications in talking toys and the like."[33] The functional categories at work here are not linguistic, nor do they resemble other ways in which the voice has been transformed into document, for example, as in the copyright of a radio show. It would, in other realms, be very difficult to get copyright on talking in a conversational way. In the material world the ownership of voice has been redefined.

Recording Chips

This category encompasses many of the most recent voice chips products. It is the existence of these products that tests the nature of the communication that we have with these technologies: do we, can we, converse with these products? The category draws from the other typologies but is distinguishable, for the most part, by the recording functionality that is raison dêtre of the product. This category includes those products that perform a more specific speech function that could not be alternatively represented by lights, beeps or visual display, i.e. perhaps they are more communicative. This category includes the products that seem to hold dialogue.

The range of products include the shower radio that reinterprets bathing as a time for productive work, an opportunity to capture notes and ideas on a voice chip, consistent with the theory that there is an ongoing expansion of the work environment into "private" life. It also includes both the recordable pen and its business card size counterpart, the memo pad. Both the pen and the pad have many versions on the market currently, and they seem to be becoming more and more populous. The "YakBak" is the parallel product for children, deploying the same technology with different graphics, and to radically different ends.

The growing popularity of this category compared to the others arouses a number of questions. Firstly, how do we understand why this category is popular? Is the popularity driven by consumers because these products are successful at what they do? And is what they do, dialogue? Or is it that the cost and portability of the technology makes it an affordable new tech symbol beyond what is attributable to their function alone? Is this is a popular category because they alone can be marketed as a work product?[34] And then conversely, why are these devices not more popular? Why is it that only a few types of products become the voice sites (i.e. pens, photoframes, memo pads are all documents of a sort, in contrast to switches or menu choices)? According to the patent literature the failure of the market place to find a need for voice capability on home appliances has discouraged the use of voice chips in other products[35] but lending the market agency for design assumptions is circular logic. This does express, however, the sentiment that many more products could have speech functionality then do.

Although miniaturization has made these products possible, the concept of embedding recording capability in products has been possible with other technologies. There has been no technical barrier to providing recording capability in cars for instance or in any of the larger products, a refrigerator for instance, certainly since the existence of cheap magnetic recording technologies. Why is it that now we want consumer products that talk to us?

It is striking that the majority of talking products on the market currently are for conversing with oneself. Although deeply narcissistic, this demonstrates a commodification of self-talk that transforms the conceptualization of the self into a subjectivity in relationship with our products. It suggest, without subtlety, that the relationship with these products is a relationship with the self. The constitution of personal and social identity by means of acquisition of goods in the market place,[36] the process of identifying products that provide the social roles we recognize and desire, can not be excluded from the consideration of the social role of products.

Where the voice chips speak
The above typologies focus on what the voice chips say rather than where they say it. However, because voice chips are distributed throughout the product landscape, where they appear (and disappear) is also interesting to examine. Although a detailed analysis could yield an interesting social geography, it is beyond the scope of a paper only intended to generate preliminary questions about why they say what they do where they do.

The automobile industry, a highly competitive, heavily patented industry that quickly incorporates cheap technical innovations (where they do not substantially alter the manufacturing process) is a place to expect the appearance of voice chips. Indeed there was early incorporation of voice chips in automobiles. A 1985 luxury car, the Nissan Maxima, came with a voice chip as a standard feature in every vehicle. The voice chip said: "your lights are on"; "your keys are in the ignition"; and "the door is ajar." There were also visual displays that marked these circumstances, yet the unfastened seatbelt warning only beeped. By 1987 you could not get a Nissan Maxima with voice chip, even on special request. In this case, the voice was silenced, but only for a time, reemerging with a very different role to play in the automobile.

By 1996, the voice chips reappeared in the alarm system of cars. Cadillacs standard alarm system uses proximity detection to warn you are too close, please move away. In this ten year period the voice shifted from notification to alarm, a trajectory from user friendly to a distinctly unfriendly position. It is also interesting to note another extension of the action/reaction voice chip logic, if not the voice itself. The current Nissan model no longer notifies that the lights have been left on, it simply turns the lights off if the keys are taken out of the ignition. The courtesy of notification has been dispensed with, as well as the need for a response from the user. The outcome of leaving the lights on is already known so the circuit will instead address that outcome. This indicates that when the results are exhaustively knowable, the need for interaction diminishes.

Of the seven patents specifically for vehicles[37] all bar one are intended for private and not public transportation. However in late 1996 voice chips began to appear in the quasi private/public vehicles of Yellow Cabs of New York. After debate about what ethnic accent[38] should be ascribed to the voice that reminded you to: "please fasten your seatbelt" and "please check for belongings that you may have left behind," a prerecorded (68k quality) voice of Placido Domingo and other celebrities won the identity contest, and since has proliferated into many well know New York characters, from sports stars, to "Sesame Street's" Elmo. The voice chip in this quasi-public sphere adopted a broadcast voice, albeit poor quality, or a micro-broadcast voice. Whether they are effective in increasing seatbelt wearing or reducing the number of items left in the cabs in any accent is less certain than the manner in which they articulate the social relations of the cab. The voice chips address only the passengers and assume that the drivers don't hear them, although it is the drivers who bear the brunt of their monotony.

Their usefulness delegates the human interaction of service and rests on the assumption that the chips are more reliable and consistent in repeating the same thing over and over, no matter the circumstance, and that the customer responds to Placido Domingo's impassive, recorded reminder more than they would a driver who may be able to bring some judgment to bear upon the situation. In the transformation of the passenger into a public audience (not unlike that of a radio station) the product or service itself is not attributed with the voice. Instead the voice becomes identified with a celebrity.

In the transportation sector alone we can see the voice chip develop from an anonymous to an identifiable voice, and from a polite notification to an alarm for deterring approach. Cars have struggled with the problem of talking to humans and seem to have exploited the non human qualities of their speech,[39] the things that the technology is better at doing, like the faithful repetition or their careful reproduction of the identity of another, rather than any particularly human attribute of their speech. It is also notably that they have not endured.

In another social sector highly saturated with electronic product, the health industry, the distribution of voice chips is almost exclusively on one side of the home/professional, expert/non-expert divide. Although in number, there are more products made for hospitals and clinics than the home market, the placement of voice chips is inversely represented. From the menstrual cycle meter to the Cardiopulmonary Resuscitation (CPR), the electronic voices seem to play the role of the health professional or "expert." In addition, the large number of products that are intended for the visually impaired, are intended for the visually impaired patients and not professionals (a demographic with more spending power); see, for example, the addition of a sound indicator to the syringe filling device for home use, which testifies that the user of this device is imagined at home, without the help of the professional for whom the product can stand in. Ironically, the most vocal equipment in this industry are the relaxation and stress reduction products, i.e. talking to yourself or being reassured and relaxed by the sounds of the ocean (see, e.g., Figure 7). The reassuring factuality of these techno-voices, focuses its attention on the lay audience.

These are preliminary observations of the voices introduced into transportation and in the health and medical areas, and are cursory at best. But they demonstrate that for the voice to make sense, the technological relationship itself needs to make sense. The speech from devices is as culturally contingent as language.

There are many other areas in which their introduction provides insight into what technological relationships make sense. Their incorporation into work products articulate the transformations and reorganization of work structure particularly into "mobile" work.[40] They speak to a cultures popular notions of where work gets done, a culture in which providing a product to take voice notes while in the shower makes sense. The voice chip

population of areas of novelty products, children toys and educational products, and of the safety, security and rescue products also maps the social relationships we engage in with our products. Conversely, where we don't find voice chips, for example in biomedical equipment for health professionals, also maps the social relationships that the technologies plays out. However, to understand the dialogue we are having with these voices requires us to also examine how we listen.

Discussion

Voices Chips as Music

The preceding categories survey what voice chips say, where it is they say it, and to whom they say it. To understand what the voice chips are saying, however, means engaging strategies for listening that may not be automatic. Products, with or without voices, are well camouflaged by what Geertz (1973) described as the dulling sense of familiarity with which ... our own ability to relate perceptively to one another is concealed from us. Modes and strategies for listening that can help us hear these voice chips can be borrowed from music. Music, unlike machines, is commonly understood as culture, or a cultural phenomena and its analysis looks very different in comparison with the analysis of technology. Perhaps the most glaring difference is the concept of improvisation, which can describe much of interaction with machines, while prevalent in theorizing music, is unusual in the analysis of human machine interaction. For our examination of voice chips aligning with music is a strategy to avoid the contests over reality, progress and rational choice that usually inform the analysis of technology and can thus provide more emphasis on the interpretative experience. Additionally, some of the voice chip products themselves that demonstrate an indifference to the distinction between speech / music, by blurring the distinction between words and beeps (see the Speech as Music category of products).

Music, like product, is also easily recognized as involved in the production of identity. That is, subcultures identify through and with music.[41] Where technological product is presented to the consumer, at what Cowan call the "consumption junction," we are at such an identity-producing site.[42] For this reason it is difficult to ascribe any one particular meaning or mode of listening to the voice chips. In the wide spectrum of musical styles available each piece of music can and does exist in widely different listening situations. This means that each listener has a variety of listening experiences and an extensive repertoire of modes of listening. The hearing person who listens to radio, TV, the cinema, goes shopping to piped music, eats in restaurants, or attends parties, has built up competence in translating and using music impressions. This ability does not result from formalized schooling, but through the everyday listening process in the soundscape of modern city. Stockfelt asserts that mass media music can be understood as something of a nonverbal lingua franca,[43] without of course denying the other more specialized musical subcultures to which we may simultaneously belong.

Listening modes are not, of course, limited to music, and nor for that matter is a musical experience limited to music. Even so, teasing out the musical modes of listening from listening modes that focus toward the sounds quality, its information carrying aspect, or other nonverbal aesthetic modes is difficult. The cultural work of using unmusical sounds as music is not uncommon, for example, Chicago's Speech Choir, John Cage's "433," the "Symphony of Sirens"[44] and the sounds created with samplers, particularly for percussive effects. At the same time the sirens, speech choirs, etc. do not lose their extra-musical meaning as they become music. Conversely, using musical sounds for nonmusical ends is the conceit of many voice chip applications.

The two products above demonstrate the confusion of musical listening vs. other modes of musical sound consumption. The Soother uses unmusical sounds for musical effect while the Funny Animal Piano using musical sounds to respond to toddler's feet. The alignment of voice chips with music has interesting implications for their linguistic claims, if they produce meaningful speech why don't they differentiate between music and speech?[45] Is it that the social position of the product determines the meaning of the sounds and utterances? Indeed if the speech they produce is linguistic, then when the voice of the alarm system warns us are we altering the meaning of the sound whether it resembles speech or siren? Or can we expand linguistic theories to accommodate all meaningful sounds that humans or machines make? These questions about how we understand the sounds that the voice chips produce, complicate the attribution of agency to these things with voices. Voice chips seem to frame sound as a prepackaged cultural product, the identity of which is located in the manufactured materiality. At the consumption junction these voices are heard in the buzz and squeal of products, but can we call it language?

Voice Chips as Speech

What do the voice chips tell us about our understanding of language? The voice chips stabilized language in material form provide a picture of our on-the-ground, in-the-market operationalization of language. Even though some voice chips use music and speech indistinguishably, the words that they say cannot be overlooked. Voice chips talk and say actual words, but how do we understand these voices as communicative resources? Are they speech acts, as defined by linguistic theorists?[46]

Speech acts[47] are used to categorize audible utterances that can be viewed as intending to communicate something, to make something happen or to get someone to do something. To construe a noise or a mark as a linguistic communication involves construing its production as a speech act (as opposed to a sound that we decide is not communicative). Categories of speech acts are given below (examples quoted from voice chips):

Commissives: speaker places him/herself under obligation to do something or carry something out, promises for example, or in a telephone system, "I will transfer you to the operator";

Declaritives: making a declaration, that brings about a new set of circumstances, when your boss declares your are fired or when the car states the lights are on; Directives: tells the listener to do something for the speaker, please close the door," "move away from the car";

Expressives: without specific function except to keep social interactions going smoothly, like "please" and "thank you," or the more expressive "I love you."

Each of these categories is performed by the voice chips examined in this paper, as are other verbs and verb phrases that are associated with the wider category of elocutionary acts: to. . . state, assert, describe, warn, remark, comment, command, promise, order, request, criticize, apologize, censure, approve, welcome, express approval, and express regret.[48]

Searle defines the "speech act" as utterances (actions) intended to have an effect on the hearer, with preconditions and effects. This has been criticized by other theorists who have pointed out that meaning is imparted by the work of an "interpretative community."[49] The limitation of speech act theory in explaining voice chips is that it ascribes the most intention to the least animate thing in the interaction. In its failure to elaborate on interpretation it provides no place for information about the significance of any particular assertion, warning, or more generally, any speech act. Voice chips amplify this problem because they can inhabit so many different situations yet repeat the same thing. Because the voice doesn't change, all flexibility in understanding to accommodate the changing circumstances needs to be accounted for by the listener's interpretation. The case of the Cadillac's alarm voice illustrates this.

In a demonstration of the Cadillac's alarm system the salesman instructed me to move away from the car and then approach the car. Despite coming as close as I could to the car the voice did not sound. On hearing no voice, the demonstrator toggled the key fob switch. I approached again and the voice sounded. In the first approach the voice chips silence was interpreted as the alarm is not working or is not on. In the second approach the voice communicated "now the alarm is on and functioning." By staying in the proximity range of the alarm system the voice answered several questions despite it repeating the same words "move away..." What is the area range in which we are detected? Will the alarm keep repeating or will it escalate its command? Although moving away from the car stopped the voice, we also came to understand the types of motions that it detected, the speed of approach, what happened when we physically shook the car, etc. The simple interaction with the car and its voice demonstrates the interpretative flexibility that transcended the directive of the words stated and how, as hearers, we respond to the voices imperatives. So in asking how we understand the significance of speech performed by the voice chip we are asking whether speech is abstractable.[50] In other words, is there a difference between talking with a voice chip and talking with something (human) with which we share capacities other than speech?

Is speech abstractable?

Speech in action, rather than in theory, is conversation. If we are to claim that we interact with voice chip speech then we need to examine the fundamental structure of conversation as the primary model for interaction.[51] One of the voice chip patents claims the rights for electronic apparatus(es) for talking in a conversational manner on different subjects, deriving simulated emotions which are reflected in utterances of the apparatus. While the other voice chip products make no explicit claim to be conversing they do claim to be "interactive."[52]

The work of Lucy Suchman may prove more appropriate to describing the interactive "speech" of voice chips. Her work focuses on the inherent uncertainty of intentional attributions in the everyday business of making sense via the conversational interaction with another machine, the photocopier. Like voice chips, she characterizes machines by the severe constraints on their access to the evidential resources on which human communication relies. She elaborates the resources for constructing shared understanding, collaboratively and in situ, rather then using an a priori system of rules for the meaningful behavior. Suchman shows that the listening process of situated language is dependent on the listener to achieve the shared understanding of successful communication. The listener attends to the speakers words and actions in order to understand. Although institutional settings can prescribe the type, distribution and content of talk, for example, cross examinations, lectured, formal debates, etc., they can all still be analyzed as modifications to conversations basic structure. Suchman characterizes interactional organization as (a) the preallocation of turns: who speaks when and what form their participation takes; (b) the prescription of the substantive content and direction of the interaction, or the agenda.[53] Thereby a system for situated communication, conversation is:

1. An organization designed to support local endogenous control over the development of topics or activities and to maximize accommodation of unforeseeable circumstances that arise; and
2. Resources for locating and remedying the communication troubles as part of its fundamental organization.

Conversation with a voice chip?
Prerecorded voices of voice chips are ill equipped to detect communication troubles, and although they are usually triggered by local inputs the content of what is said does not change. They will repeat the same thing or a set of prerecorded phrases over the indefinite range of unpredictable circumstances. While localizing control they, for the most part, do not localize the direction of speech.

The type of application that seems closer to Suchman's characterization are the products that include "dialogue chips." These chips quite literally hand over control of the content of talk to the listener, fulfilling Suchmans characterization of conversational interaction in this respect. The listener literally controls the speaker and sets up a relationship with the device. Further, the dialogue chip products uses the turn taking of conversations collaboration, not as the alternation of contained segments of talk in which the speaker determines the units boundaries, but in the manner illustrated by the joint production of single sentence.[54] The "turn taking system for conversation demonstrates how a system for communication that accommodates any participants, under any circumstances, may be systematic and orderly, while it must be essentially ad hoc."[55]

Therefore, the response to voice chips, like the applause at the end of a play, is not a response to the final line uttered, or the fact that it just stopped. "the relevance of an action...is conditional on any identifiable prior action of event, insofar as the previous action can be tied to the current actions immediate local environment." The conditional relevance does not allow us to predict from an action a response but only to project that what comes next will be a response, and retrospectively, to take that status as a cue to how what come next should be heard. The interpretability therefore relies on "liberal application of post hoc ergo prompter hoc."[56] The response that a listener can have to the voice of the train defect annunciation system is not only a response to the words uttered by the product. It will also involves a complex series of judgments that includes assessments of the information available and how to integrate into what else the listener can know of the event at hand.

The understanding of talking products does not come so much from the words at what is popularly conceived as the human-machine interface, but beyond this. The voice is a voice embedded in a network of local control, sequential ordering, interactional organization and intentional attribution.

But it is the recordable chips with which we can have a dialogue with ourselves that best demonstrate this. These products literally frame the understanding that we are talking with ourselves through our products. While dialogue is conversation with another agent, one whom is there somehow, monologue is characterized as written speech, inner speech or rehearsed speech. Dialogue implies immediate unpremeditated utterances, whereas monologues are written speech lacking situational and expressive support that therefore require more explicit language. Questioning the abstraction of speech in voice chips does not demonstrate that speech is uniquely human. On the contrary, the stabilized voices of hardware based speech are subject to reinterpretation and rediscovers the listeners capacity, not the speakers incapacity. It may simply be viewed as a distinction between dialogue and monologue, neither of which are more or less human. Because we inhabit both sides of a dialogue we can understand the voice chips position and compensate so as to perform dialogue with ourselves.

From Voice Chips to Speech Recognition
This paper has so far developed the unique position of voice chips products, differentiating them from the background noise of contemporary culture and other technological configurations that deliver speech. These hardware bound voices are not broadcast and have no stable identity. The survey of what the voice chips say produces typologies that suggest further investigations of how we understand and use these voices, where they appear and what their voices mean. The short product life cycle of the consumer electronic devices they inhabit position these products as the E-coli of socio-technical relations and can demonstrate the formation of product identities, products voices, in the shifting understandings of machine interaction. The appearance of voice chips in some types of products and not others, some social sectors and not others is open to further investigation. Detailing these would reveal the voice chips oral history of the process by which the very ephemeral social device of the speech becomes stabilized and entered into systems of exchange.

Before concluding I introduce a complimentary examination of speech recognition chip sets, around which there is much more recent product development activity. While the voice chips applications seemed to have peaked around 1997, the equivalent low power, distributed speech recognition function may be just beginning. Watching their development and deployment carefully, asking now that we can talk to our products, what do we say? may allow us to hear the social scripts they presume. However, because we are more self conscious about speaking than listening this may be an instrument through which to observe our own roles in socio-technical interaction. In order to prime this investigation, and because speech recognition chip sets are not yet (and may never be) widely available, the author hosted a competition to survey a range of applications. The competition was advertised on a large mailing list (12,000), the Viridian list owned and carefully managed by science fiction writer Bruce Sterling. The list is a forum for discussing technological futures with an emphasis on addressing environmental problems. Entrants were asked to propose a speech recognition interfaces to an existing product (the prize was a voice note taker and the prototyping of the proposed device), just under three hundred designs were

submitted and are available on the Web site <www.cat.nyu/neologue>. While these entries cannot be claimed to represent the conceptions of human computer interaction distillated by the social forces of the market, manufacturing and advertising we see crystallized by the voice chips, they can be treated as evidence of technological desires, expectations and hopes, that may or may not be observable in the market. Now that we can talk to our device, what do we say? The most striking feature the competition entries demonstrated is the explicit intention to effect social change with technological change. This may or may not be peculiar to this list (which can be tested by hosting a similar competition in other contexts) however, this is consistent with a popular techno-determinism that attributes social change to technological change and under-represents the dominant forces of product innovation that can be attributed to sustaining and continuing a corporate entity.[57] This also contradicts other popular understandings and lay rationalizations that new products arise to address preexisting social needs or profit opportunities, follow fashion or to optimize existing applications.

We can briefly summarize the trends illustrated by the proposed products and product interfaces[58] (a longer analysis in Jeremijenko forthcoming) which is predominantly the desire for social and individual envisioning and regulation. This is apart from the ultimate (and theatrical) control fantasies that this particular type of interface engages (e.g. on saying "showtime" the lights come dim and the television and VCR turn on),[59] or the suggestions that substituted buttons without explicating the word, e.g. dispensing with the TV remote,[60] but not explicating what words exactly to use. Entries that do not explore what happens in the translation from finger-button to voice-button and the social (and observable) spectacle this makes do not render the socio-technical relationship this investigation is trying to identify. There were also the applications that were similar to the voice chips — with a similar use of speech/buzz interchangeably in the applications that called attention to itself, e.g. the cookie container that recognizes childrens footsteps to trigger singing, or the TV remote that calls out polo when it hears marco.[61] The self-observation, regulation and control, take on and moral, physical, emotional, and consumption monitoring and regulation in such forms as: a wallet that recognized words and dispensed consumption regulation advice;[62] a pocket device that recognizes "now what am I supposed to do?" and responds "with a gentle reminder to adhere to the users selected ethical set"[63] (regulation of consumption); coffee maker that recognizes "good morning", "when you respond the chip analyzes your tone of voice" [for sluggishness].... "adjusts the "strength of the coffee... " (automating the physical regulation on which Starbucks has successfully capitalized); or the more extreme circumvention of your own self judgment, in monitoring bloodflow and detecting stress the "device whispers "relax", dims the lights a bit, and releases soothing aromatherapy";[64] or the very opposite of an alarm clock which would be a device that on hearing "why am I still up?" "should cause every light and entertainment system in my house to shut off for four hours." An example of the self-observation, was a voice triggered "nocturnologue"[65] which would record any sleeptalking. These devices to regulate the self, toward social synchronization presumably, do not necessarily imagine the devices as "companion" and attribute it a more social performance, although there is a small subset that do. This subset of entries realize the "technology-should-be-more-human-like" expectation, that reflects a similar school of Human Computer Interface (HCI) designers working towards adaptive interfaces, that can recognize and respond to different emotive state[66] as an explicit strategy to be "user-friendly." The best example is a comedic sidekick (Jerry Lewis), ready with smart rejoinders on recognizing phrases and built into the watch[67] (when it hears "nice hair," the device says "cha cha cha"). This functionality would have to be described as reinforcing social performance.[68] This seems both similar to other identification relationships (cars, furniture, home) and different insomuch as it is directly inserted into the conversation.

The promise of emotive interfaces to recognize and respond to how you are feeling,[69] if these imagined interfaces are any evidence, was demonstrated and expressed in words that describe an ambivalence, even resentment, of technological relationships: for example being able to say "shut up" to your television set[70] or to your telephone[71] (not "turn off," not "close/finish" or other ending command). Clearly, this complicates the sort of understanding we can develop about a persons relationship to a product from the purchase of it. And this is of course the predominant form of "feedback" that companies and designers get about products. These voices make audible a strongly polarized ambivalence. There was no suggestion of saying "I love my TV" to turn it on, that is otherwise invisible.

Another device was proposed for automated prayers, triggered by saying "pray for me," it is customizable to different religious "preferences,"[72] took this further. Prayers suggested included excerpts from Psalm 23 to "Cynical hipster types [who] might want their in-dash prayer boxes to recite William S. Burroughs Thanksgiving Prayer (Thanks for Indians, to provide a modicum of challenge and danger... thanks for a nation of finks... etc.) and some guilty white liberals (some Viridians, even) might want theirs to apologize for driving around in a vehicle spewing noxious fumes into the atmosphere."[73] This is more than an interface to recognize and respond appropriately to user emotional states; actually the entertainment is in delegating the emotionality or at least religiosity itself to the device. This impulse is replicated in the delegation of care, social niceties and other arational and non-calculative tasks to the computational devices. For instance, a speech recognition chip that recognizes the sound of flatulence and politely apologizes to the room,[74] relieving the responsibility of any one person to bear the embarrassment; another entry, as an extension of Tamagachi-like automation care, suggested using a voice recognition chip to train the parrot to speak.[75] There were actually several other entries exploring information technology for animals which seems to be evidence against a voice interface imagined as "humanizing" the computer, and more a demonstration that the ready treatment of animal noises as recognizable sounds imagines these as functionally equivalent in every way to English words. Speech recognition, reinterpreted as sound recognition.

Finally, and perhaps the most interesting or novel constellation of projects, are the designs that use the opportunity to script interactions as a form of propaganda, propaganda that is distributed (enacted) beyond traditional and corporate monopolized media channels. The portable idealogue was suggested to play the role (and potentially look like) the soapbox.[76] The BackTalk is a portable billboard for one's car. It is triggered by the use of simple trigger words and suggested deep set LEDs to display specifically to the driver behind a message of "thanks for letting me in," "baby on board," or presumably any other bumper sticker expression. This is intended to influence others and begins to populate this category of the regulation (or at least influencing) of others. This has very direct and explicit forms: many in fact directed at those currently not well socialized cell phones, which, for example, cut out if they hear you say "yeah, I am on the cell phone," "yeah, I am in the village," "Dude"[77];or monitor for swear words[78] and other efforts to silence loud or otherwise "inappropriately" private voices in public spaces; to quite many suggestions directed at rendering massified phenomena. This social observation impulse is illustrated by an entry that is a museum display designed to collect responses (what the entry called clichés) so that "will grow as an open ended accretion or demonstration of the clichés uttered by thousands, tens of thousands, millions of art consumers", and that this collection itself is the spectacle. The museum exhibit is rethought as an instrument for the collection of comments and the desire is to see the massified phenomena. This is the desire for seeing a social spectacle that is repeated often and I would like to argue is a recurrent theme in the networked context of information technology. Another suggestion was the "crowd morality barnacle" which is a device intended to influence mass behavior, in this example in a riot. This CMD is intended for distribution throughout a crowd and will respond to key riot phrases, e.g. "smash," with "be careful," "burn" with "it might explode"; or "get them" with "where are the children."[79] This is a different conception of regulation than the examples that illustrated the control of self. To effect self control the designs went beyond turning electronic devices off or regulating the self with insistent and unrelenting reminders, e.g. correcting a habit of speech or cutting the "umms" out of the story, to quite novel punishment. These punitives enacted on the self included squirting water in your ear, triggering electric shocks, dribbling water down ones leg. There were few viable designs that offered a simple reward rather than punishment. To effect the social body, while there were no physical punitives, the reward seems to have been the social behavior itself, or at least the evidence of it (as in the spectacle of clichés).

The final category to describe is one that relies on the double entendre of words, simultaneously using several meanings of the words. This was explored by some of the entrants and is important to understand that it demonstrates that the speech interface cannot be understood as making the machine more human. Rather, it is clearly exploiting the different parsing, context sensitivity and repeatability of human vs machine models of cognition. For example, to trigger the discrete recording of conversations one entry describes a recorder that is triggered by "what's up amigo." This deployment of an unusual (relative to the user and context of use - i.e. no one else is likely to say it) filler is used to initiate conversation and direct attention of the people being addressed but simultaneously being used for instrumental purposes as the on button. Likewise the "don't hurt me, just don't hurt me" cell-phone/gps position locator/911 dialer proposal[80] uses a self defense phrase to dial for help without alerting the presumed attacker, who is presumed to hear the plea on face value – second guessing a reasonable or usual response in a threatening situation. The interaction here is the user being able to employ simultaneous meanings of the words they use. And that clearly the speech chip is being used so that the words used to interact with the machine, are understood to be different from the speech used to interact with humans.

It is also notable that there were categories of speech not explored by these interfaces. Consider the linguistic communication defined as a performative. A performative, such as "I do," is a highly codified and stabilized utterance that communicates a future commitment or social contract.[81] Because it is a stabilized social technique it would be technically pragmatic, the problem of unlimited variation of phrasing is solved, were not subject for speech recognition chips. The absence of designs to address this sort of statement is curious, and worth further investigation.

These categories of interaction demonstrated by this brief survey of voice chips are not discontinuous or radically different from other contemporary consumer technologies. The observation of self (or ones own property) is embodied in the consumer video camera market, and surveillance systems; self regulation has extended from alarm clocks once a day to alarming cell phones carried with you and ready for all alarming occasions; handhelds directly regulating sleep and activity, to vcr/tivo to capture, regulate (in order to extend) and meter out media program consumption; social observation is also embodied by surveillance systems but although surveillance looms large in the popular imagination it has not been used to see or envision the mass or each other. The problem of seeing the social body has remained an architectural problem, solved by spectacles of plaza, and malls — public and quasi-public places. What the voice chips most clearly demonstrate is that it is this area in which there seems to be the most interest — literally being able to see the massified behavior. The traditional broadcast (e.g. television) media had very little interest in rendering the public to itself, and as such the rise of phone in, and reality television genres suggest that even in the context of high-production value broadcast media there is a cultural appetite to "see" each other, no matter how contrived. The collaborative filtering models, such as popularized by the Amazon people-who-bought-this-book button show each others behavior, to make it the shared experience— to see where others have been. Like the micro-casting of speech

recognition triggered rear window car display, we see this desire expressed through the car — and the cars peculiar access to the public space of freeways. This is a public space where the rules of communication between and amongst people are highly constrained (cf: plaza). This is not the interactive experience of the self with the self, or the self with the machine, but the machine as a proxy for interacting with the social. This is a peculiar and interesting way to think about human machine interaction.

Conclusion

The interactions we hear with the voice chips do not disambiguate the buzzes and beeps used by speechless machines, but the speech recognition products do reinforce that we use speech for machines and speech for humans differently, and simultaneously. The other applications also re-imagine how we understand their functions. The products discussed do not exploit the mechanistic, logical and fully controllable functions of machines but treats them as complicated multifarious social actors. There is a clearly stated desire to enlist these new technologies and product interfaces to promote explicit desired social transformations. We also here the ambivalent relationship we have with and for our current technological devices.

This paper has explored why listening to voice chips and speech recognition chips might give us a way to examine human machine interaction in situ. Much real complexity of social and technical interactions is lost in the tradition of examining this within controlled laboratory context, and ethnographic analysis can be too rich. However the theoretical perspective that has developed from the ethnographic insights, that privileges the improvisational nature of real world applications, enable us to focus on how speech and turn taking is used to coordination of the interaction between machines and humans.

This initial analysis is presented in order to set up some the preliminary ideas and interpretation, so that as (if) the speech recognition chips become more widely distributed we can tune-in to this particular historical moment and hear what it is we expect, want and bring to our human machine interaction. There are few instruments that give us this viewpoint. Listening to our daily interactions with products can work to contest and complicate the dominant methods used to describe technological trends and patterns of product innovation: demographically driven and massified market research and the capture of consumption behaviors at point of purchase. The examination of the speech recognition applications give unique access to the assumptions, expectations and the imaginative work of products and the interactions they script.

Further examinations of voice chip and speech recognition products and patents can extend what has only begun. Firstly, in understanding how voice chips abstract speech we can examine what we understand interaction to be and hence how we design and frame interactions in products of daily use, reproducing our understanding of human technical relations. The products make obvious the design assumptions with which they are built, but further investigation of the details of their use will help to elaborate how these micro-interactions perform and realize actual social roles and social structures. A detailed use analysis of any one of the products can provide further insight into this sort of investigation. The voice chips raise other questions too. Because they slice through many social and economic sectors but are still a manageable population of products, they can be used to illustrate the iterative and continuous process of technical change that is intimately involved in a technologies sociality, in contrast to the radical discontinuities of technological change through discovery and paradigm shifts.[82] They realize a recombinant model of technological change. Furthermore, for the same reasons they can be used to examine the changing social position of these products in relation to the configuration of power and work relations,[83] and the transformations of the market groups and users that these products presume. Finally, in the tradition of Turkles examination of children understanding of their interactive machines, childrens products with voice chips can illustrate what child care roles we delegate to machines, and articulate clearly the hardwired (per hardware not neurons) expression of consumption identity of children. For these reasons this paper marks the beginning of a project to collect an ongoing database of products with voices or speech recognition that appear on the market, or receive patents.[84] As a longer archive of product voices may prove a valuable resource for the examination of changing socio technical relations, even in the event of the products falling silent and voice chips and speech recognition being abandoned altogether.

The voices of the products reflect back the voices and interactions we have projected and programmed into them, reflecting them back for our reinterpretation. Therefore, as the title of this paper suggests, a mode of interaction we have with the consumer products that exist and are imagined at the time of this paper, is a dialogue with a monologue. By literally listening to what hardware has to say, and what we say to it, we may better ground our assumptions of interaction in reflexive reinterpretation. Furthermore, we can see from this examination that scripts of human machine interactions are used to extend the predictability of individuals and coordinate their interactions, but that there is an opportunity and expectation that this gives us a method to hear and understand these massified interactions, and see these technologies as voice and ears of the social body.

References

1. L. Althusser. (1971). *Lenin and philosphy and other essays*. New York: London Monthly Review Press.
2. While most communication theorists account for the social world, building a framework for understanding communication is often at odds with accounting for the diversity of possible experiences of language and the modulation of each social position. Austins work that looks at not how a language is composed but what it does, from where it does it. See Austin, J. (1980). *How to do things with words*. Oxford: Oxford University Press; or Volosinov, V. (1973). *Marxism and the philosophy of language*. New York: Seminar Press.
3. Quoted from the North American Patent literature.
4. Pacific Bell voice mail system 1996, 1997, and AT&T automated customer help.
5. Benveniste, E. *The nature of pronouns problems* (1956) showed how linguistic categories not only allow human subjects to refer to themselves but actually create the parameters of human self-consciousness. " Ego is he who says ego. That is where we see the foundation of subjectivity which is determined by the linguistic status of person. Consciousness of self is only possible if it is experienced by contrast. I use I only when I am speaking to someone who will be a you in my address." p. 225 The linguistic category such as "I" relies wholly on the identity of the speaker for its meaning.
6. Latour, B. and J. Johnson. (1988). *Mixing humans and nonhumans together: The sociology of the door closer, social problems*, Vol. 35, 298-310; Callon, M. *Four models for the cynamics of science*. In Sheila Jasanoff, Gerald E. Markle, James C. Petersen and Trevor Pinch (eds). (1995). *Handbook of science and technology studies*. Thousand Oaks, CA, London & New Dehli: Sage Publications, 29-63.
7. Callon, M. and J. Law. (1982). On interests and their transformations: Enrollment and counter-enrollment. *Social Studies of Science*, Vol 12, 615-25.
8. Latour published the book *Science in action* (Cambridge, MA: Harvard University Press, 1987) in 1987, while in Dallas, June 11 1978, Texas Instruments Incorporated announced the launch of its speech synthesis monolithic integrated circuit in the new talking learning aid, SPEAK &SPELL(tm). The speech synthesizer IC accurately reproduced human speech from stored (a capacity of 200 seconds in dynamic ROM) or transmitted digital data, in a chip fabricated using the same process as that of the TI calculator MOS ICs.
9. See M. Patons forthcoming Social Studies of Science paper for a detailed examination of the initial construction of the virtues and values of the phonograph recording technology.
10. Minneman, S. (1991). *The social construction of engineering reality*. (PhD dissertation, Stanford University.
11. Hallmark card first included voice chips in their cards in 1988. Five years later they introduced a recordable card on which you could record your own voice.
12. Nissan Maxima 1986.
13. Barbie said three things when she was given a voice in late 1980s: "Meet me at the Mall," "Math is Hard," and "I like school, don't you?"
14. Machina R , a San Francisco based company, had on the market in 1997 several talking pens or "Pencorders," talking keyring, several talking photoframes and many "Cardcorders," including "Autonotes."
15. Saccahrin is claimed to be the first product to be parasite marketed, i.e. "this product uses saccharin."
16. Turkle, S. (1984). *The second self*. New York: Simon and Schuster, 31.
17. A complete list of the collected products and patents is attach in the appendix. A full list is available at http://cdr.stanford.edu/~njj/vcprods. This is being updated constantly.
18. Patent # 4863385 Sept 5 1989.
19. Patent # 5313557 May 17 1994.
20. Patent # 5014301 May 7 1991.
21. Patent # 4812968 Mar 14 1989.
22. Patent # 4951032 Aug 21 1990.
23. Patent # 4863385 Sept 5 1989.
24. Patent # 5315285 May 24 1994.
25. Patent # 4987402 Jan 22 1991.
26. Patent # 5117217 May 26 1992.
27. Patent # 5254805 Oct 19 1993.
28. Patent # 5275285 Jan 4 1994.
29. Patent # 5481904 Jan 9 1996.
30. Patent # 5555286 Sept 10 1996.
31. Turkle op. cit. note 16 demonstrates how children enter into social relationships with their computers and computer games in which they thinking of it as alive and get competitive, angry, they scold it, and even want revenge on it. She finds that they respond to the rationality of the computer by valuing in themselves what is most unlike it. That is, she raises the concern that they define themselves in opposition to the computer, dichotomizing their feeling and their thinking.
32. Patent # 5413516 May 9 1995.
33. Patent # 5029214 July 2 1991.
34. Work and the products of work can be shown to take on meaning that transcend their use value in commodity capitalism see Willis, S. (1991). *Primer for daily life*. New York: Routledge.
35. Patent # 5140632 Aug 18 1992. Telephone having voice capability adapter.
36. Shields, R. ed. (1992). *Lifestyle shopping: The subject of consumption*. New York: Routledge.
37. Within the patent literature what appeared in relation to transportation were: 5555286 Cellular phone based automatic emergency vessel/vehicle location system: translates a verbal rendition of latitude and longitude to cell phone; 5509853 Automobile interior ventilator with voice activation: which queries the driver when door closes and gives menu options; 5508685 vehicle and device adapted to revive a fatigued driver: a voice reminder combined with spray device; 5428512 Sidelighting arrangement and method: voice warning of impending obstacle; 5045838 Method and system for protecting automotive appliances against theft; 5278763 Navigation Aids(presumably for application in transportation); 4491290 Train defect detecting and annunciation system.
38. See *The New York Times* discussion.
39. This is in contrast to the popular depiction of cars with voices on mainstream television, in programs such as "My Mama was a Car" or "Night Rider" on CBS, the voice was used to lend the car personality.
40. Zuboff, S. (1984). In the age of the smart machine: *The future of work and power*. New York: Basic Books. In particular, see The abstraction of industrial work, 58.
41. See Fabbri, F. (1981). A theory of musical genres: Two applications. In *Popular music perspectives*, eds. David Horn and Phillip Tagg. Gothenburg and Exter: International Association for the Study of Popular Music.
42. See Oswald, L. (1996). The place and space of consumption in a material world. *Design issues*, Vol. 12 (1), who describes the site for purchasing product as the staging of the subject in consumer culture.
43. Stockfelt supports her work with Tagg and Clarida studies on listeners responses to film and television title themes that demonstrate common competence of adequately understanding and contextually placing different musical structures. That listeners for the most part understand the musical semiotic content in such situations in similar ways, across cultural areas that are more dissimilar.See also Tagg, P. (1979) Kojak, *50 seconds of television music: Toward the analysis of affect in popular music*.
44. The symphony of the "Sirens" first performed in 1923, Arseni Avraamov.
45. In particular, the products that use speech and music interchangeably: the childrens applications, bells and whistles substitute for spoken encouragement, or the alarm systems that will use vocal warnings or sirens sounds, the pen patent #4812068.
46. To relate the voice chip to the socio-linguistic universe and its emphasis on the place of language within it, interprets the social system as a semiotic, and stresses the systematic aspects of it. We cannot simply assume that the concept of a system itself and the concept of function (of language) within that system is the most appropriate starting point. However this assumption underlies most of the guidelines developed for computational models of speech and is thus appropriate for discussion of the voice chip.

47 Austin, J.L. (1980). *How to do things with words*. Oxford: Oxford University Press. The general point of which was not to look at how language is composed but what it does.

48 Searle uses this list to introduce his paper: Searle, J. (1972). What is a speech act? In P.P. Giglioli ed. *Language and social context*. Harmondworth: Penguin.

49 Stanley Fishs essay How to do things with austin and searle. In *Is there a text in this class? The Authority of Interpretative Communities*. Cambridge, MA: Harvard University Press, 1980, 244 analyses Coriolanus as a speech-act play. When Coriolanus responds to his banishment from Rome by stating a defiant "I banish you" the discrepancy in the illocutionary force in both the performatives of banishment is obvious. Rome, embodying the power of the state and community vs. Coriolanus sincere wish to banish Rome, i.e. his intentionality, is illustrative.

50 Broadcast voices and prerecorded voices although abstracted onto technologies still belong to an identity, however it is the combined sense of abstraction that connotes the identity of the voice as that of the car. This could be interpreted alternatively as an abstracted voice of authority performed by the car or the abstraction of the car itself.

51 "If certain stable forms appear to emerge or recur in talk, they should be understood as an orderliness wrested by the participants from interactional contingency, rather than as automatic products of standardized plans. Form, one might say, is also the distillate of action and interaction, not only its blueprint. If that is so m then the description of forms of behavior, forms of discourse...included, has to include interaction among their constitutive domains, and not just as the stage on which scripts written in the mind are played out," (E. Schegloff). Discourse as an interactional achievement: some uses of "uh huh" and other things that come between sentences. In Georgetown University Round table on language and linguistics: Analyzing discourse text and talk. D.Tannen, ed. (1982). Washington, DC: Georgetown University Press, 73.

52 Patent # 4517412 the card actuated telecommunication network is an example of this. "Local processor 11 controls a voice chip 15 coupled to telephone set 10 which interacts with the caller during the card verification process."

53 L Suchman, L. (1987). *Plans and situated action: The problem of human machine communication*. Cambridge: Cambridge University Press. Suchman explains that this interpolation of verbal nuances and the coherence that the structure represents is actually achieved moment by moment, as a local, collaboratively, sequential accomplishment. The actual enactment of the interaction is an essentially local production, accomplished collaboratively in real time rather than born whole out of the speakers intent of cognitive plan, 68-98.

54 Ibid.,81, 125. Suchman uses the example of the joint production of a single sentence to demonstrate the fluid division of labor in speaking. and listening.

55 Ibid., 78

56 Ibid., 83.

57 product innovation for corporate continuity — assessing the life expectancy of corporate products.

58 A longer analysis in Jeremijenko forthcoming.

59 A.M.Dixon@shu.ac.uk MikeyMoneyMinder.

60 Ibid.

61 zoeluna@bellsouth.net.

62 Ibid.

63 Ibid.

64 Afrench@iss.net (Andre French).

65 nocturnologue.

66 mit emotive interaces again - brooks; this is in contrast to the Shneiderman et al work that argues that this works against control.

67 Wristwatch sidekick.

68 This is a version of the gestural value of handheld and portable devices identified and described in a study involving the ethnographic examination of filmic depictions of the use of handhelds. Jeremijenko (1992) XEROX PARC internal publication.

69 For example – MIT media lab – emotive recognizing facial expression.

70 vbar@comp.cz (Vaclav Barta).

71 butler@comp-lib.org (Michael Butler).

72 The peculiarity of referring to a religious preference as if it were another consumption category – that religious and addictive behavior is subject to the same economic characterization....?

73 jon@lasser.org (J. Lasser).

74 butler@comp-lib.org (Michael M. Butler).

75 wapel@tc.cac.edu.eg.

76 xiane@entech.com.

77 spiff@bway.net.

78 monitoring for swear words.

79 zoeluna@bellsouth.net (Dave Whitlock).

80 zoeluna@bellsouth.net.

81 see Judith Bulter (1996).

82 Dosi, G. (1985). *Technological paradigms and technological trajectories research policy* 11 :1982) 147-162; and Clark, K. The Interaction of design hierarchies and market concepts in technological cvolution. In *Research policy* 14, 235-251.

83 See for instance Zuboff op. cit. note 40.

84 This list is available at cat.nyu.edu/neologues and is being updated constantly. It includes images and product literature and when possible an audio file recording of the voices.