Unexpected, Unremarkable, and Ambivalent
OR
How the Universal Whistling Machine Activates Language Remainders

Marc Böhlen
University at Buffalo
marcbohlen@acm.org

JT Rinker
University at Buffalo
jtrinker@buffalo.edu

ABSTRACT
In this paper, we describe a situated machine that allows us to explore the idea of the language remainder from an unusual perspective.

Keywords
Language processing, philosophy of language, languages without words, low bandwidth direct communication, interface design and belief, animal-human-machine interaction

1. INTRODUCTION
One often takes for granted the many ways of knowing that are needed in order to make meaningful statements about the world. Logically correct statements are comparatively easy to formulate because the rules of logic are known and formalized; sensible and meaningful statements, as in statements of common sense, are much harder to specify. Computer systems have been struggling since their inception with this problem. What are the requirements for formalizing the automation of making sense? How can one test if "sense" has been made? In this case, people are still the best arbitrators. For this reason, the automation of making sense includes the capacity to interact with people on their terms, to reason about their input on their terms, and to share the output again on their terms. On the input side alone, the problem is exceedingly difficult. The most obvious natural input modality, speech, is difficult to automate. Despite decades of research, no speech recognition system is capable of graceful and robust recognition when dealing with unrestricted vocabularies and untrained speakers [1, 2]. But there is no need for pessimism. One day this problem may be solved. And incremental improvements in the machine learning community may just bring about the complex kinds of data representations we will be hard pressed not to call meaningful.

2. The Language Remainder
While general-purpose meaning may be achieved using current computational approaches, it is questionable whether fine-tuned variations of meaning, layered, contradictory, situation-specific, and temporally limited meaning can be captured. Fuzzy aspects of language such as innuendo defy formal linguistic descriptions and are not even modeled in computational models of language that seek to represent "all aspects of communication" [3]. Assuming that the making of meaning will remain intimately linked to language, it makes sense to ask how language should be represented in machines capable of making sense. Language is more than a database of words and rules by which these words can be combined. Languages are not static, and not fully describable through the grammatical rules that constrain them, however refined such rules may be. Many philosophers of linguistics, semioticians, and writers have pointed this out. Lévi Strauss reflected on the counterintuitive robustness of systems of semantic impoverishment [4]. Eco revealed the follies of those who have tried to uncover, rediscover or create a perfect language, failing because of their inflexibility, utilizing rigid rules [5]. Maturana and Varela spoke of language as a process, not as a static collection of words and rules. They coined the term 'languaging' to better represent the richness and complexities created in the exchange between cognition, language and language use [6]. Lecercle proposed the term 'remainder' as a formal entry into the levels below, above, and adjacent to strait-laced meaning covered by linguists' version of language [7]. For Lecercle, the remainder is the fallout from the intended use of language. It is the essence of poetry and metaphor, but also of miscommunication, word play, and double-entendre. It is the fuzziness and leakage of meaning amongst words.

But how could one possibly attempt to include the language remainder in computational systems? Is it at all possible, given that the rigor of linguistics seems even tighter in the limited corpora of texts, the defined rules and intelligent but blind numerical clustering methods underlying computational linguistics? In order to prevent varied and flavored meaning and language remainders from being filtered out of computation, it might be worthwhile to investigate varied and less structured forms of knowing, unorthodox methods of input, and unexpected flavors of output. Easier said than done. This is not only a difficult problem, but also a poorly defined one. How can one even begin to formulate such issues as tasks, let alone make them computationally tractable? We offer no general solution to this problem. However, we suggest an alternate approach towards the problem. Would it be possible to reduce the complexity of language to a more manageable subset, albeit one that still allows instances of language remainders to exist? Rather than creating a machine that is conceived with hardwired

First published at COSIGN-2004,
14 – 16 September 2004, University of Split (Croatia),

knowledge of a fully structured language, including vocabulary and grammatical system, would it be possible to create a device that is only ‘primed’ for language? Is the ability to perceive and imitate a limited bandwidth of data that is mutually suggestive by machine and user as communication, a precursor to language? Can meaning arise in such a situation? In this context we offer the following thought experiment:

Imagine walking down a corridor, past an elevator, lost in thought. You hear a whistle. You stop and search in curiosity or disdain for the person seeking your attention. Nothing. You notice an intercom-like device embedded in the wall. Again you hear a whistle. You walk to the device, stare at it. Another whistle. You whistle back. The device whistles again, in a different fashion. You respond, and realize now that you have engaged in a whistle exchange with a machine.

3. Accessing the Language Remainder

We have designed, built, and tested such a device and call it the Universal Whistling Machine (U.W.M.). It senses the presence of moving creatures in its vicinity and attracts them with a signature whistle. Given a response whistle, U.W.M. counters with its own composition, based on a time-frequency analysis of the original. The following paragraphs describe the contexts in which we place this device, and how we succeed and fail to introduce the slippery language remainder into a machine.

Figure 1. U.W.M. in standalone mode.

4. Unremarkable Forms of Expression

How can whistling include the above-mentioned potential of language remainders? Is it too limited a form of expression? Maybe not. Whistling, like other forms of low-bandwidth communication – grunting, coughing, laughing, and humming - are phoneme-less modes of communication. One might consider such modalities to be extreme cases of homophony. The fact that one word can have many meanings has led to interesting debates on the limits of language, such as, - are there more things than words in the universe? If we are structurally confined to the homophony condition, why not make a virtue of the vice? Whistling is extreme homophony.

There are practical reasons that make such low-bandwidth forms of expression interesting. Technically, it is much easier to analyze whistles than it is to analyze spoken language. The frequency spectrum of whistles is far simpler than the frequency spectrum of other vocal emissions. A low-bandwidth, high-contrast signature is typical of whistles and contrasts sharply with voice signals. Because of this, whistles can be readily differentiated from singing [8]. Furthermore, whistling is a low-entropy mode of information encoding, and easier to separate from a noisy background than speech.

But technical details should not guide this discussion. Is there a merit -beyond delivering a tractable signal-processing problem- that makes such low-bandwidth exchanges desirable and interesting?

We believe the answer is yes. Whistling is an underexposed and unremarkable area of expression: It is raw and direct. Whistling is immediate, code and content at once. Whistling is pleasure, admiration, warning, unfiltered desire, cipher, and protest. Emmet Till, a young man of color, was lynched in 1955 after wolf-whistling in the presence of a white woman. Intuitively understood, whistling is transcultural communication below the radar of social etiquette. The idea that language is material seems uncontested in such low-bandwidth expression – here the body speaks on its own terms. Coughing and humming bring body fluids to our lips. If language is material, then it is in the forms of low-bandwidth expression that this becomes most apparent. Language, as material, unfolds. Everything said is said by someone [6], and every utterance is uttered through a mouth. In whistling and humming we revert back to less articulated states of communication. Whistling is simple, unassuming, and highly unremarkable. Yet, as De Landa observes, it is in the unremarkable where human universals are to be found [9].

4.1 Agency via Semantic Impoverishment: Languages Without Words

There are numerous examples of human communication systems based entirely on whistling. This phenomenon was widely reported during the late 1970s in linguistics’ circles [10]. Two of the better-known whistling languages are “el Silbo”, practiced on the Isla de la Gomera, one of the Canary Islands off the coast of Morocco, and the whistled language of Kuskoy, a remote village by the Black Sea in Turkey that has only recently been connected to the telephone grid. In La Gomera, the skill is still being passed on to youngsters today. These two whistling languages contain a subset of the spoken language with which they coexist.

But why do such whistling systems even exist? Some linguists have speculated that they originate from the need for secrecy and robustness. High-pitched strong whistles travel farther than the spoken word. In La Gomera, the maximum distance that a whistled message can travel has been reported to reach 10km. Skilled whistlers are said to be capable of producing whistles of 130dB (measured 1 meter in front of the whistler [11]).

Whistled languages are generally “reduced” languages, in the sense that not everything that can be expressed in speech can be expressed by whistling. However, they are far closer to languages than to a code or to simple signalization systems. They are speech-like and carry the vocabulary, the grammar, and in many cases the phonology of the local language they have emerged from, especially at the level of prosody.

Whistling occurs across languages and cultures. Whistling is a primitive communication in most human languages. Whistling is a kind of time travel to a less articulated state. Whistling carries the potential for song, pleasure, and secretive message. All people have the capacity to whistle, though many do not whistle well. Since it lacks phonemes, whistling is a pre-language language. Unlike other forms of low-bandwidth encoding, such as hand clapping, whistling is nuanced and offers the advantage of a rudimentary dialogue. Within the laws of digital signal processing, human utterances and machine-based audio signals are more similar than different. Whistling is much closer to the phoneme-less signal primitives compatible with digital machinery than to the domain of spoken language. As such it offers itself an Esperanto of communication not only across language boundaries, but also across species boundaries.

But how should one represent an artificial whistle? And how can one interface with it?

5. Interfaces beyond Navigation and Control

Artifacts are inscribed with involuntary meaning. Artifacts have signatures. They reveal their signature in the materials and techniques used to make them and in the logic and worldview under which they were conceived. In this sense, any artifact is a statement about the world. The computer interface as artifact is no exception and reveals a particular belief system about how we are expected to engage with machines. The usual distance between a human and a computer is arm's length with a keyboard as the mediating device. Robust but rigid, this configuration has come to define the human computer interface. Attempts to make this interface more 'natural' tend to include direct input modalities, i.e., modalities that require no adaptation by the human. Speech and gesture recognition suggest that a person could communicate with a machine as if the machine were human. Despite the fact that the implementation is troublesome, the idea seems appealing. But even when the problem is solved, we may be no closer to language remainder compatible exchanges with machines than before. The problem is not just technical.

In 'Behind the Blip,' Matt Fuller analyzes Harun Farocki's "I Thought I Was Seeing Convicts" [12]. This video consists of video clips, footage from inside the California state prison, Corcoran. The notorious reputation of the Corcoran penitentiary is based on the rumor that prison guards were alleged to have set up fights between members of rival gangs in the exercise yard. Guards then watched the yard via video surveillance and bet on the outcome of the fights. Fuller points out that the "interface" here is not where one might expect it to be. The screen in front of the guards is the site where the visual cues were available, but the structural interface that made the scenario possible is built into the 24-7 surveillance system, the image storage capacity of the video system, the prison architecture, the command structure of prison officers, and the range and caliber of the guns used to enforce order [13].

Interfaces extend beyond the immediate point of sensory exchange between the human and the machine. Interfaces are only meeting points, rhizome-like, between other dependent and informing mechanisms. Interfaces are made up of objects and the use of these objects over time by people with various needs and desires. The effectiveness of an interface is a function of what it allows as well as what it disallows. As with language, there is an interface remainder, something that falls between the cracks, something the user manual does not mention.

6. Situation as Interface

U.W.M. recognizes when people are approaching by analyzing the data from a built in video camera. The machine then waits until a person who has entered its field of perception moves outside of it before emitting a whistle.

We use a blatantly direct method of attention seeking. Moreover, it is a method that is decidedly outside of standard human-machine interface conception. Machines do not initiate exchange and machines do not whistle at people. Here, the device latches onto a particular form of human-human 'interface' for its agency. The awkwardness of this situation can provide the activation energy necessary to ponder alternatives to established interface models. To whom should we complain when a machine harasses us? What then is the nature of this kind of harassment – the whistle itself or the imitation of desire? And the people that decide to turn around and approach the machine are not 'users'. Our machine offers no instructions or visual cues on how to 'use' it. U.W.M. has no interface in the traditional sense. Only those who choose to whistle back at the machine will be able to extract something from it. U.W.M. is not picky. Whether one produces a fine-tuned melody or a rough and graceless whistled sound is of no significance for the machine. U.W.M. works with the whistle it gets.

Figure 2. U.W.M. folded into a wall

In its deployed state U.W.M. is embedded in a wall, and the situation it initiates becomes its interface. This anti-interface creates a window of ambivalence in which the casual pedestrian is offered an unusual experience. Being whistled at by a human might generate feelings of anger, disgust, or pleasure. Being whistled at by a machine forces humans to recognize their preconceptions about the role of the machine vis-à-vis the human. U.W.M. has no goal state other than to generate such a situation, such a discontinuity of expectation, from which an experience akin to the pleasure of the language remainder can result.

In order to make such events likely, we are installing foldable and portable U.W.M.s in quiet, low-traffic spaces of exchange and transition. Restrooms, corridors, and elevator halls (Fig. 2) are examples of transition spaces that suit this experiment. They are multi-purpose spaces, home to a variety of "services" not inscribed into the formal architecture. For instance, people tend to linger around while waiting for an elevator. Some people may wander down corridors lost in thought. The underdetermined nature of activities typical of such locations creates a temporary semantic void that is well suited for the experience U.W.M. offers.

Figure 3. European Robin (*Erithacus rubecula*)

7. Meaning Beyond Human Understanding

As whistled sound producers, machines, humans, and animals share a common denominator. Whistling and song we share with animals. Mammals and birds have the means for making songs and whistles. Ducks, robins, loons, and starlings whistle. White whales and bottlenose dolphins [14] whistle under water. Just as we carry physical remnants of our bodily evolution in us, we have the capacity for whistling. When we whistle, we acknowledge the plane of being underneath phonetically articulated language that we share with other species. Beyond alternatives to computer interfaces, U.W.M. also offers the potential for a new approach to human-animal communication.

U.W.M. is capable of imitating certain bird whistles as easily as it can synthesize human whistles. (See our examples collection for a synthesized robin whistle [15].) Would a songbird imitate a whistle from a whistling machine just as readily as it imitates whistles from birds and certain human-generated sounds? Could this lead to new modes of human-machine-animal exchanges? Can we find alternatives to exchanging information beyond the species boundary? And when we do, which rules shall we apply? [16]. Interaction research to date has centered on human-machine exchange. The worn-out narrative of subduing our surroundings is unconsciously but consistently inscribed in the computer interaction program. How can we open ourselves to the otherness of the animal state, so utterly unlike our own that Agamben calls it 'unopen' [17]. Can the language remainder, made partly accessible by machines like U.W.M., help us open the unopen? St. Francis derived great admiration from his ability to speak to the animals.

8. The Remainder Persists

How do we speak of the unspeakable? Where is a 'right' place to begin an investigation of the informal? Sometimes detours are the direct path. That is what Paul Miller observes on his tour through the art and science of rhythm. Miller starts with sound and works through literature, the visual arts and philosophy, only to land back at mystery of music, informed by the art of the DJ [18]. It was all in the mix from the beginning. Before language there was gesture, then utterance. The past is phonemeless. Pure signal. Before the word forms on your lips, air is pressed through your mouth. Today, the word rules. Tomorrow, the wordless utterance of the machine that is so far beyond human it is animal again, awaits us.

9. Technical Notes

The U.W.M. project is coded in C++ and PD. U.W.M senses people passing by via a low-cost, IEEE1394 enabled CCD camera/sensor at video rate with a public domain camera driver [19]. The data from this video stream is parsed by standard and custom-made machine vision routines. Oblong objects traversing the camera's field of view at the speed of a casual pedestrian trigger the device into whistle mode. The machine waits until a person who has entered its field of perception moves outside of it before a whistle is produced.

Given a whistled response to this invitation, U.W.M. will in turn whistle back at an individual. Sound capture occurs through a USB enabled noise-reducing microphone. Signal sampling occurs at 44.1 kHz. In order to prevent U.W.M. from erratically responding to all kinds of sounds and noises, the audio input is analyzed for its 'whistleness'. We have developed a robust decision mechanism that allows us to react specifically to whistled input. Incoming signals are passed through an FFT based pitch tracker, and analyzed on two passes. Data outside of the three-sigma boundary are discarded as outliers. The filtered data is then reprocessed for the standard deviation, curtosis, median and bandwidth. Select boolean conditionals based on these parameters allow us to exclude most unwanted signals. The device is able to reject background noise, including music (pop, hip-hop, strings and high wind instruments and song from both male and female voices). Cries from young children, however, can elude the current detection mechanism.

9.1 On the Whistleness of Whistles

Other researchers have previously investigated the properties of whistles [20]. Here the aim is different and the return value is different. This machine speaks as spoken to. It whistles back at an input whistle. The schematic below gives an overview of the algorithm we use for U.W.M.'s artificial whistles.

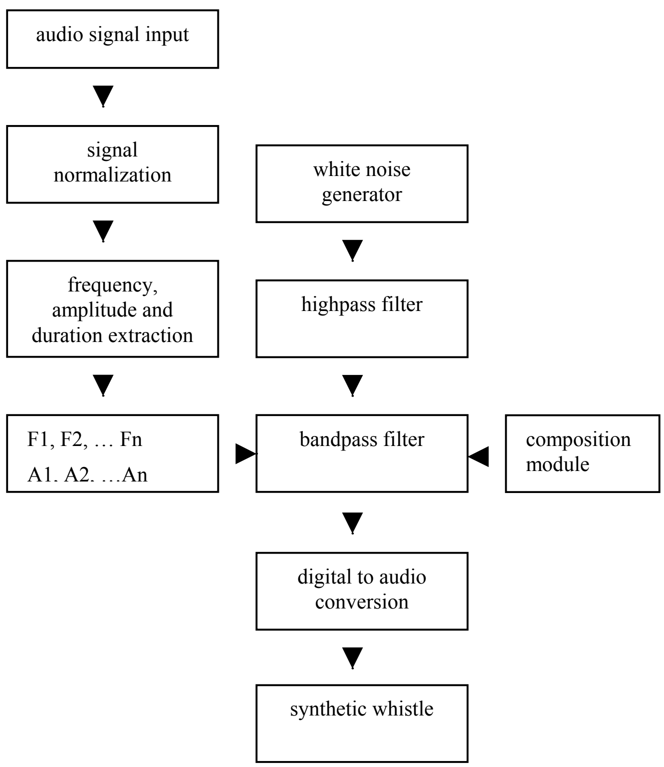

Figure 4. Schematic diagram of the main audio processing components in U.W.M.

The whistle synthesizer is based on the basic spectral characteristics of a human whistle. Most human whistles exhibit a fundamental frequency with very few harmonics, (often only one or two) as well as a band of high frequency noise. Here, we create a whistle through a process of subtractive synthesis. We use white noise as a signal generator for the whistle synthesizer. The noise is passed through a pair of filters in series. The first is a single-pole high pass filter. This filter eliminates most low frequency noise. The second filter is a band-pass filter, which passes a sinusoid at a specified center frequency and attenuates all other frequencies. The center frequency is the pitch for the whistle, and the "Q" (quality factor, similar to bandwidth) of the filter is set proportional to the center frequency. Data from the pitch-tracking device drives the whistle synthesizer. However, the prototype machine can also generate random whistles by supplying its own center frequency and amplitude data to the synthesizer.

9.2 From Synthesis to Transformation

To generate responses to perceived input whistles, raw data is collected from the pitch tracker as frequency and amplitude pairs. High threshold gates on amplitude content as well as other algorithms, including interpolation in the pitch domain and peak detection in the amplitude domain, help smooth out areas in which the pitch tracker fails. This is most notably the case at the ends of whistled pitches and between attacks of whistles where pitched signals are not present. These smoothing operations allow us to extract from the data stream only those elements essential to the whistle itself. In addition to being able to imitate input whistles, we have also created numerous forms of whistle transformations. For instance, adding a fixed pitch interval to the pitch data creates a transpositional transformation. This results in a response whistle that is higher or lower in pitch than the input whistle. Contours of the input whistle can be increased, decreased or inverted to give a semblance of the shape of the input whistle while varying the pitch. Time transformations read the data at a rate different from the capture rate. This creates responses that are slower or faster, and independent of pitch and amplitude and their subsequent transformations. Tempo rubato is created by randomly changing the read-time interval between each index of the pitch and amplitude arrays, thus speeding up some portions of the response whistle while slowing down other portions. The data can be read backwards as well as forwards, essentially reversing the input whistle. All of these transformations can be applied in parallel providing a wide palette of responses that are all based on a user's input whistle. We make use of Miller Puckette's PD environment [21] for these operations.

10. ACKNOWLEDGMENTS

This text is dedicated to the memory of Hope Kurtz. The mockingbird knows.

11. REFERENCES

[1] R. Cole, J. Mariani, H. Uszkoreith, A.Zenen, V. Zue, "Survey of the State of the Art in Human Language Technology", Center for Spoken Language Understanding CSLU, Carnegie Mellon, 1995.

[2] M. Maybury, Language Technology: A Survey of the State of the Art Language Resources: Multimodal Language Resources, MITRE Corporation, 2002.

[3] R. Hausser, "Foundations of Computational Linguistics: Man–Machine Communication in Natural Language", Springer Verlag, 1999.

[4] C. Lévi-Strauss, "The Savage Mind (La Pensée Sauvage)", University of Chicago Press, 1967/1962.

[5] U. Eco, "Serendipities, Language and Lunacy", Harcourt Brace & Company, 1999.

[6] U. Maturana, F. Varela, "The Tree of Knowledge", New Science Library, 1987.

[7] J. Lecercle, "The Violence of Language", Routeledge 1990.

[8] D. Gerhard, "Computationally measurable differences between speech and song", PhD Thesis, Department of Computer Science, Simon Fraser University, 2003.

[9] M. De Landa, "A Thousand Years of Nonlinear History", Zone Books, 1997.

[10] R. Busnel and A. Classe, "Whistled Languages", Berlin, Springer Verlag, 1976.

[11] J. Meyer, “Comparison typologique des langages sifflés: processus cognitifs”, PhD thesis, University of Lyons, (synopsis on the web: http://www.ddl.ish-lyon.cnrs.fr/)

[12] H. Farocki, “I Thought I Was Seeing Convicts”, with text “Controlling Observation” in: Harun Farocki, Nachdruck/Imprint. Texte/Writings, Lukas and Sternberg, 2001.

[13] M. Fuller, “Behind the Blip”, Autonomedia, 2003.

[14] L. Rendell and H. Whitehead, “Culture in whales and dolphins”, Behavioral and Brain Sciences (2001), 24, 309–pp. 382.

[15] M. Böhlen, “More Robots in Cages”, Space&Robotics2000, 4th International Conference on Robotics for Challenging Situations and Environments, p. 206 - 212, Albuquerque, New Mexico, 2000.

[16] www.buffalo.edu/~mrbohlen/uwm.html

[17] G. Agamben, “L’uomo e l’animale” (The Open, Man and Animal), Bollati Boringhieri / Stanford Press 2002/2004.

[18] P. Miller, aka DJ Spooky That Subliminal Kid, Rhythm Science, MIT Press, 2004.

[19] The CMU 1394 Digital Camera Driver is maintained by the Robotics Institute at Carnegie Mellon University. See:

http://www-2.cs.cmu.edu/~iwan/1394/index.html

[20] L. Prechelt, R. Typke, “An interface for melody input “, ACM Transactions on Computer-Human Interaction (TOCHI), Volume 8 Issue 2, 2001.

[21] M. Puckette, "Pure Data: another integrated computer music environment." Proceedings, Second Intercollege Computer Music Concerts, Tachikawa, Japan, pp. 37-41, 1996.