LEONARDO: "The Word---Voice, Language and Technology", (in print 12/2005). Please do not quote from version -

Experiments with Whistling Machines

Marc Böhlen and JT Rinker

The MediaRobotics Lab, Department of Media Study, University at Buffalo
231 Center for the Arts
Buffalo, NY, 14260

Abstract

This text describes a series of experiments performed with a machine capable of synthesizing human whistles and canary song. We call this device The Universal Whistling Machine (U.W.M). It senses the presence of living 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.
A previous paper [1] described the whistle synthesis and U.W.M. with regards to language representation. Here we give a broad overview of all issues U.W.M touches upon, including animal-machine interaction.

1 Introduction

Thoughts need language so that they can be shared with others. In order to have agency, our thoughts need to be, in some manner, compatible with the thoughts and ideas of others. One path into a compatibility of thought might be the notion of common sense. Indeed, we often take for granted the amount of effort required to make sense, and the amount of work it takes by others to understand our thoughts and spoken words. This has become especially apparent in the decades of research into computational representations of language and thought. The requirements for formalizing the automation of making sense are far from obvious. There is no procedure by which one can test if "sense" has been made. Logically correct statements are comparatively easy to formulate and verify because the rules of logic are known and formalized; sensible and meaningful statements, as in statements of common sense, are much harder to specify. The computational expert systems hailed only a few decades ago failed because, for one reason, the role of common sense was underestimated. Since then ambitious research projects have been devoted to the issue of including common sense in computational knowledge representation [2]. But to date the best experts of common sense are still human beings. And so inquiries into automated sense making must include the capacity to interact with people on their terms, to reason about their utterances on their terms, and to share the responses again on their terms. For computers, utterances are inputs, and the most obvious natural input modality is speech. Despite years of research, no speech recognition system is capable of graceful and robust recognition of even a single language, let alone multiple languages simultaneously [3, 4]. This is particularly obvious when speech recognition systems are confronted with unrestricted vocabularies and untrained speaker voices. Under this speech recognition impasse, the task of automating common sense becomes even more challenging .
No doubt, this will change in the future. Promising improvements from language modeling and machine learning may just bring about the complex kinds of data representations we will be hard pressed not to call useful and meaningful. But there is more to language than pure utility.

2. A Machine Primed for Language

While general-purpose meaning may be achieved using current computational approaches, it is questionable whether fine-tuned variations of meaning; layered, contradictory, situated, culturally specific, and temporally limited meaning can be captured and created. Slippery aspects of language, such as innuendo, defy formal linguistic descriptions and are not represented in computational models of language that seek to represent “all aspects of communication” [5]. 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. Wittgenstein described vividly how only the use of words and not their definitions define their actual meaning [6]. 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 [7]. 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 [8]. 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 [9]. For Lecercle, the remainder is the fallout from the intended use of language. It is the essence of poetry and metaphor, word play, and double-entendre. It is the fuzziness and leakage of meaning amongst words. It is both the source of meaning between the lines and miss-communication. The potential of such a process-oriented understanding of language as languaging has also been confirmed by computational linguists investigating language as an adaptive system with emergent properties [10, 11].

But how could one attempt to include the language remainder in computational systems? Computational linguistics with limited corpora of texts, defined rules and intelligent but blind numerical clustering methods does not, to date, appear to hold much promise in this regard. 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 representing knowledge and language. It might be worthwhile to experiment with alternate ways of knowing, unorthodox methods of accessing machines, and unexpected types of responses. If the richness of the language remainder is unattainable through direct representation of language in the machine, it might be accessible by the representation of other features of communication outside of spoken language. This is easier said than done. We offer no general solution to this problem. However, we suggest a replacement problem for the original question. 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 knowledge of a single fully structured language, including vocabulary and grammatical system, it might be possible to create a device that is only “primed” for language, but ready for languaging. Might the ability to perceive and create a limited bandwidth of data that is mutually suggestive by machine and user as communication be a precursor to languaging? In this context we offer the following thought experiment:

> Imagine walking down a corridor lost in thought. You hear a whistle. You stop and search in curiosity or disdain for the person seeking your attention. You notice an intercom-like device embedded in the wall. Again you hear a whistle. You walk towards 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.

We have designed, built and tested just such a machine. We call it the Universal Whistling Machine (U.W.M.) It senses the presence of moving creatures with its built-in camera and attracts people with a signature whistle [Figure 1a, b]. 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 how we succeed and fail to access the slippery language remainder through a variety of experiments with whistling machines.

Fig. 1a,b

3 When Code is Content

Whistling, like grunting, coughing, laughing and humming are low-bandwidth, phoneme-less modes of expression. Technically, it is much easier to analyze whistles than it is to analyze spoken language [12]. 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 [Figure 2a, b]. Because of this, whistles can be readily differentiated from singing [13]. Furthermore, whistling is a low-entropy mode of information encoding, and easier to separate from a noisy background than speech.

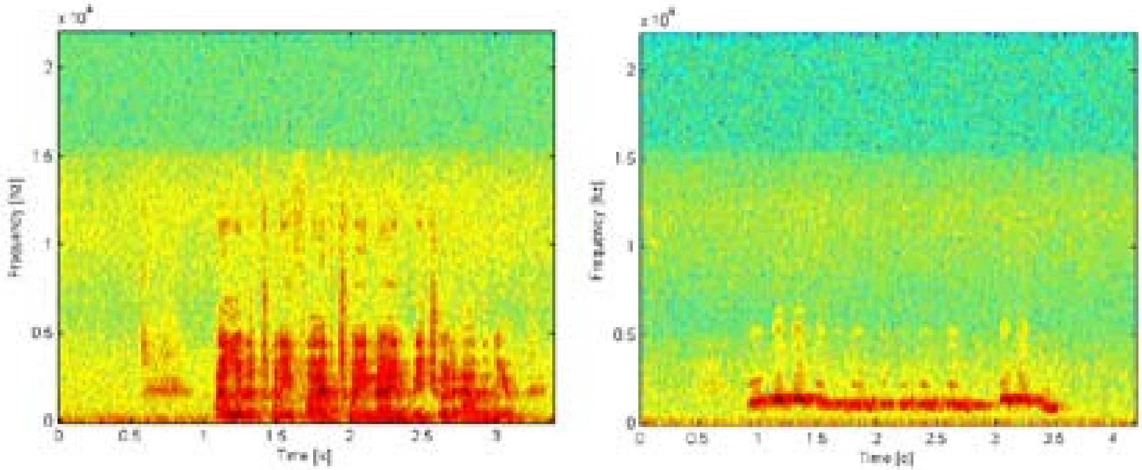

Fig.2a, b

There are reasons beyond delivering a tractable signal-processing problem that make such low-bandwidth utterances noteworthy. Whistling is an underexposed and unremarkable area of expression. Whistling is immediate, both 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. Raw, direct and intuitively understood, whistling is transcultural communication below the radar of social etiquette. The idea that language is material is uncontestable 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 [8], and every utterance is uttered through a mouth in motion. 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 [14].

4 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 [15]. 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.

Some linguists have speculated that whistled languages originate from the need for secrecy and robustness. High-pitched and forceful 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 [16]). 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 language than to a code or to a simple signalization system. 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, it is a kind of time travel to a less articulated state. Whistling carries the potential for song, pleasure, and secretive message. Whistling is a primitive communication in most human languages. All people have the capacity to whistle, though many do not whistle well. Lacking 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 as an Esperanto of communication not only across language boundaries, but also, potentially, across species boundaries.

5 Synthetic Whistles

U.W.M. whistles in response to audio signals recognized as whistles. The whistle synthesizer that makes use of this input 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, and use white noise as a signal generator. The noise is passed through a pair of filters in series. The first is a single-pole high pass filter that 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 device can also generate random whistles by supplying its own center frequency and amplitude data to the synthesizer. The diagram below [Figure 3] describes the main components of the whistle synthesis.

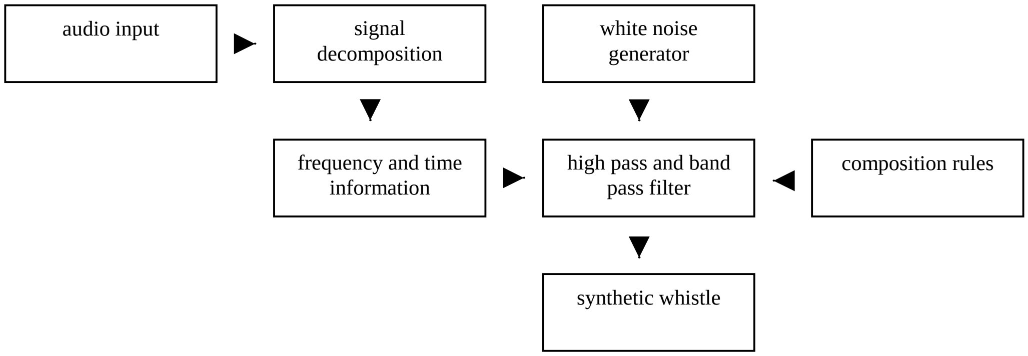

Fig. 3

To generate responses to perceived input whistles, raw data is collected from the pitch tracker [17] as frequency and amplitude pairs. High threshold gates on pitch content and low 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 are performed on the data before it is sent on to the transformation process. It allows 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 either 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, reversing the input whistle. All of these transformations can be applied in parallel, providing a wide combinatorial palette of responses that are all based on an input whistle.

Whether one produces a fine-tuned melody or a rough and graceless whistled sound is of no significance for the machine. U.W.M. responds to a given input, regardless of its level of perfection. Should one whistle a second time, the machine will counter again, differently. The longer one interacts with the machine, the more varied and complex its responses become, all the while only playing with what it receives; waiting, always ready for more.

6 Immunity to the Spoken Word

It is important for U.W.M. to be selective regarding the many sounds it is exposed to. Just as we chose to react to some types of sounds and images and ignore most everything else, U.W.M. is very sensitive to some signals and immune to most others. This signal processing based selectiveness has a correspondence in social interactions. Negation is a statement in its own right, and “NO” is a very strong word. Indeed, refusing to respond to someone is the height of disrespect.

In order to guide U.W.M to respond only to whistles and prevent it from responding to other sounds and noises, the audio input is analyzed for its whistleness in a pre-processing stage. We have developed a robust decision mechanism based on real time statistical data analysis methods that allows us to react specifically to whistled input. Speaking to the machine will generate no result; U.W.M is immune to the spoken word. It reacts only on its own terms, and those are the terms of the whistle. The whistle recognition is achieved by a two pass analysis of the results delivered from the above described Fast Fourier based pitch tracker. Data outside of a three-sigma boundary are discarded as outliers in the first pass. This filtered data is then reprocessed for the standard deviation, kurtosis, median and bandwidth.

Select boolean conditional statements based on these parameters allow us to exclude most unwanted signals in real time. U.W.M. 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. There are further signal processing techniques one can use to improve selective rejection, however, we choose to let the machine have this particular weak spot.

7 Situating a Whistling Machine

U.W.M. has no interface in the traditional sense and offers no instructions or visual cues on how it should be "used". Only those who choose to whistle at the machine will be able to extract something from it. 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 people to recognize their preconceptions about the role of the machine vis-à-vis the human. In order to make such direct confrontations more likely we have installed portable U.W.M.s in quiet, low-traffic spaces of exchange and transition. Restrooms, corridors, and elevator halls 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.
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 [Figure 4a, b].

Fig. 4a (screen shot), b

U.W.M. borrows a blatantly direct method of attention seeking. Moreover, the method is decidedly outside of the standard human-machine interface conception. Machines do not initiate exchange and machines do not whistle at people. Machines are invariably either designed to react to our actions, or to initiate standardized reactions, as do traffic lights at an intersection. But when machines become deliberative and leave the dictum of direct efficiency behind them, everything changes. When a machine purposefully and believably borrows from the complex field of social behavior patterning reserved for human beings, it transgresses an invisible boundary and lands in uncharted territory. And when it borrows from behaviors frowned upon by some, it claims a socially sensitive niche for itself. In contrast to the Social Robotics agenda calling for machines that behave like well-adapted people do [18, 19], this machine decidedly misbehaves. But the real cause of discomfort – the whistle itself or the imitation of desire on the part of the machine – eludes direct detection. Stung by a lewd whistle, an angry whistle victim might sue for unspecified damages, but would be hard pressed to identify the perpetrator let alone specify the said damages.

8 Two Whistling Machines

In the course of our experiments with the whistling machine an expansion to the original scheme suggested itself. Two whistling machines together might create something one machine alone could not [Figure 5a, b]. Given that the whistling machine is a listening machine once it stops whistling, a single initial whistle between two identical machines running the same code should in principle be able to start an endless exchange. We tested this and received rather disappointing results. Since the machines continuously altered the initial signal, and added an occasional glitch in the process, the original whistle rapidly morphed to a signal outside the bounds accepted by the whistle recognition system. As is the case in other forms of emergent behavior, the robustness of shared whistling is highly dependent on initial boundary conditions. Once we simplified the initial whistle tune and altered the composition algorithm in only one of the machines to mimic the other machine’s whistle, while leaving the composition module in the other machine unchanged, the whistle exchanges continued for extended periods of time. This generates continuously evolving whistle patterns, opening the door to further multi-machine experiments. For our current purposes, the whistle machine population is limited to two. When no one is present to whistle with, the two whistling machines, having nothing better to do, initiate their own exchange. The moment somebody enters the field of view of either machine, they stop.

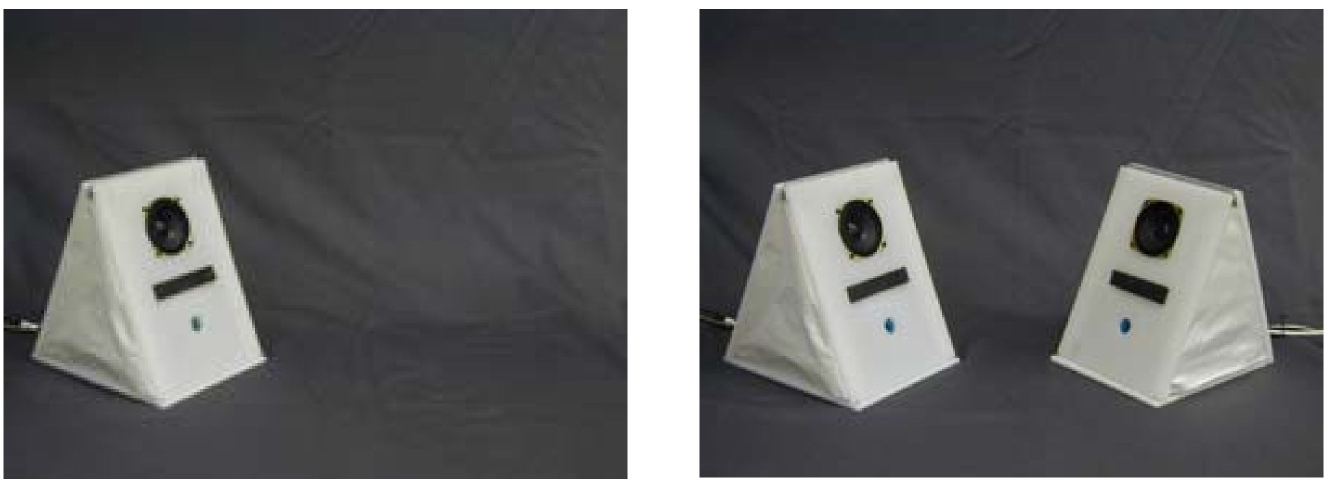

Fig. 5a,b

9 A Canary and a Whistling Machine

As whistled sound producers, machines, humans, and animals share a common denominator. Whistling and song we share with animals. Mammals and birds have means for making songs and whistles. Ducks, robins, loons, and starlings whistle. White whales and bottlenose dolphins [20] 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. offers an opportunity to reconsider the idea of machine-animal communication.

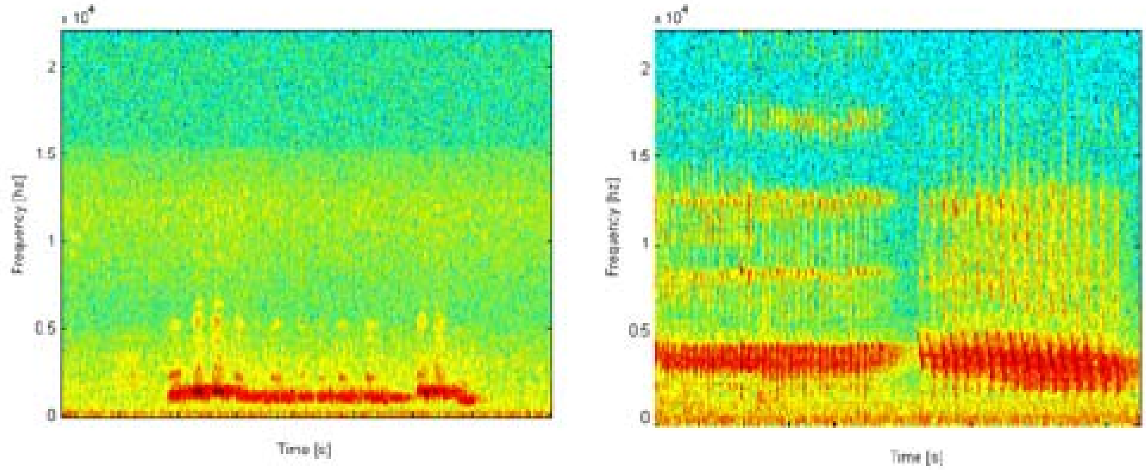

Fig. 6a, b

In order to test the feasibility of our desire to have U.W.M. communicate with other species, we chose an obvious candidate, the yellow canary (*serinus canarius domesticus*) as a U.W.M. companion, and created a new synthesis engine to decompose and recompose its song. In humans, whistles are produced by pressing air through the larynx into the mouth and out through puckered lips. Similar to the human larynx, a bird's vocal organ – or syrinx – consists of folds of tissue in the passage that connects the lungs to the throat. But canary song is succinctly different from human whistles. Most notable are the intense short bursts of energy, some lasting only 15ms [21], in which a canary emits its song [Figure 6a, b]. While the exact mechanisms of sound production in the syrinx are still a matter of debate amongst biologists, many agree that sound is primarily created through oscillation of the lateral labia—tissue folds, which open and close the air passage from the bronchi to the trachea [22]. Our canary engine does not include a full model of the bird's syrinx. However, we do model the tonal qualities of canary song.

Fig. 7

The canary engine replaces the white noise generator [Figure 3] that works well in creating the imperfect windy whistle typical of humans with a pure sinusoid. Furthermore, the frequency decomposition and recombination in the canary engine is modified for the short energy bursts typical of bird song. The repertoire of U.W.M.'s song elements comprises segments of the canaries own song repertoire, segments of human generated whistles and purely synthetic tones reminiscent of synthesizer music. While the bird showed little excitement towards both human and synthetic whistles, it clearly reacted with agitation and hopped closer to the audio source when confronted with its own sound patterns. This is consistent with observations from the ornithology community [23], where song matching has been found to be an aggressive signal in many bird species. With this in mind, we altered the interaction and composition algorithm to specifically avoid direct imitation and placed our machine together with the canary in a quiet room over a period of several days [Figure 7]. U.W.M. was programmed to be silent and wait for the bird to sing – these songs could be as long as thirty seconds- select a small subset of the signal, as short as two seconds, and alter it through frequency shifting (type-matching) or temporal inversion. While the results sounded different than the original to our ears, our canary seemed to "recognize" the origin of the signal and reacted with agitation, more so when the signal was transposed up than when it was transposed down.
In humans, mimicry creates an odd form of pleasure by insinuating some form of "understanding" on the part of the machine. With the canary, we observed no such foolish behavior.

Fig. 8

While our canary experiments generated mostly qualitative results, and are probably of little use for ornithologists, we like to think such useless tests can help us re-imagine mediation and interaction with machines. If nothing else it makes clear that our current models of machines as pets, servants or slaves are only a subset of a much larger and richer domain of information processing paradigms. In the future, we will be able to exchange data beyond the species boundary. And when we do, new rules will have to guide us [24]. 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. U.W.M. offers a different story, told by a machine, full of signal, imperfect but patient and ready for the "unopen" [25] animal state, so different from our own.

10 Temporary Access to the Language Remainder

Whether animal or human, those who choose to play with U.W.M. receive in exchange a most unusual machine experience. We have shown U.W.M. in a variety of venues and found that many visitors were swayed to spend time with the machine. Many people tried to test the limits of the variational capacity of the machine, or teased the device with chirps, clicks and whispered secrets. In their attempts to fool the system, they pushed language to the limits, mixed speech with song, and made their bodies produce strange sounds. Some visitors returned again and again to toy with the machine, and react to its reactions until the question of who is playing with whom became blurred. And each time the machine found a whistle sufficiently different from its own signature whistles, it stored it to disk, amassing over time many examples of whistles. Together these sounds form a database of past experiences, residing quietly in memory until activated.

Video documentation of U.W.M. is available here:www.realtechsupport.org/new_work/uwm.html

11 References

[1] M. Böhlen, JT Rinker, "Unexpected, Unremarkable, and Ambivalent OR How The Universal Whistling Machine Activates Language Remainders", Computational Semiotics in New Media and Games (COSIGN2004), University of Split, Croatia

[2] http://www.opencyc.org/

[3] 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.

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

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

[6] U. Eco, "Serendipities, Language and Lunacy", Trans. William Weaver. Columbia University Press, 1998.

[7] L. Wittgenstein, "Philosophische Untersuchungen". Suhrkamp, Frankfurt am Main, 1953.

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

[9] J. Lecercle, "The Violence of Language". Routledge, London, 1990.

[10] L. Steels, " Perceptually grounded meaning creation". In M. Tokoro (Ed.), Proceedings of the Second International Conference on MultiAgent Systems, ICMAS-96 (pp. 338–344). Menlo Park, CA: AAAI Press, 1996

[11] S. Kirby, "Function, selection and innateness: The emergence of language universals". Oxford University Press, Oxford, 1999..

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

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

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

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

[16] 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/)

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

[18] K. Dautenhahn,, "Design spaces and niche spaces of believable social robots", in: Proc. Intl. Wksp. Robot and Hum. Interactive Commun., 2002.

[19] T. Fong, I. Nourbakhsh,, and K. Dautenhahn.," A Survey of Socially Interactive Robots", Tech. Rep. CMU-RI-TR-02-29, Rob. Inst., CMU, 2002.

[20] L. Rendell and H. Whitehead, "Culture in whales and dolphins", Behavioral and Brain Sciences, 24 (2), pp. 309–382, 2001.

[21] T. Gardner,G. Cecchi, M. Magnasco, et al., "Simple motor gestures for birdsongs" Phys. Rev. Lett. 87, 208101, 2001.

[22] R. Suthers, S. Zollinger , "Producing song - The vocal apparatus", Behavioral Neurobiology of Birdsong, in: Annals of the New York Academy of Sciences, 1016: 109-129, 2004.

[23] S. Vehrencamp, "Handicap, Index, and Conventional Signal Elements of Bird Song" In: Animal Signals: Signalling and Signal Design in Animal Communication, pp. 277-300. Ed. by Y. Espmark, T. Amundsen, G. Rosenqvist. Tapir Academic Press, Trondheim, Norway, 2000.
[24] 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.
[25] G. Agamben, "L'uomo e l'animale" (The Open, Man and Animal), Bollati Boringhieri / Stanford Press 2002/2004.
[26] 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

12 Technical details

The whistling machines are made of polyethylene, aluminum, and electronics with a lining of white silk. Sound capture occurs through a USB enabled noise-reducing microphone array, sampled at 44.1 kHz. The algorithms are coded in C++ and PD. When in action, three programs operate concurrently. The first module, the whistle recognition engine, evaluates the audio input and finds instances of whistles in the data stream. The second module is dedicated to synthesis and transformation of the recognized whistles, and the last module is dedicated to checking if people are in the vicinity of the machine. This last module uses a low-cost, IEEE1394 enabled CCD camera/sensor at video rate and a public domain camera driver [26]. 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.

13 Acknowledgments

This project has been supported in part by the Julian Park Publication Fund at the University at Buffalo.

14 Figure captions

Fig. 1a, b:	Two different reactions to U.W.M. (Los Angeles 2004)
Fig. 2a, b:	Spectrogram of a spoken utterance (a) vs. a whistled utterance (b) by the same speaker
Fig. 3:	Schematic diagram of the whistle synthesis engine
Fig. 4a, b:	Figure capture from the internal video camera (a), and a visitor responding to U.W.M. (b)
Fig. 5a, b:	A single U.W.M. (a), joined by a second one (b)
Fig 6a, b:	Spectrogram of a human whistle (a) vs. a canary song (b)
Fig. 7:	A yellow canary with a whistling machine.
Fig. 8:	Advanced Perception, Carnegie Mellon University, MS/MFA thesis Marc Böhlen, 1999. Three Rhode Island Red hens were held in a spacious cage together with a mobile robot for 60 days. The robot was programmed to share the space with the animals and to not infringe on their habits and movements. See reference 24.

15 Glossary of terms

Fourier Transform

A transformation of a function into sinusoidal basis functions, i.e. as a sum or integral of sinusoidal functions multiplied by some coefficients.

Fast Fourier Transform:

An efficient algorithm to compute the discrete Fourier transform (DFT) and its inverse.

statistical data analysis

Data collection and interpretation according to rules of statistics

entropy
: A measure of energy or disorder in a system.

frequency spectrum
: All frequency components together given by the decomposition of a function.

white noise
: A signal with a flat frequency spectrum, having equal power in any band, at any center frequency, over a defined bandwidth.

bandwidth
: In analogue systems bandwidth is referred to as the width of a signal, usually measured in Hertz

spectrogram
: A three-dimensional plot of the energy of the frequency content of a signal as it changes over time.

signal filtering (high pass, band pass, low pass)
: The removal of certain frequency components of a signal. High pass filters remove low frequency content, Low pass filters remove high frequency content, and band pass filters remove frequencies components above and below set values.

kurtosis
: A parameter that describes the shape of a random variable's probability distribution, graphically this corresponds to the degree of peakedness of a distribution.

subtractive synthesis
: A technique used in some synthesizers to create musical timbres by filtering simple waveforms generated by oscillators.

tempo rubato
: Slight deviations from strict tempo, shortening one note or lengthening another, yet without drastically altering the overall tempo of a composition.

pitch tracking
: The continuous calculation and interpretation of frequency components of a signal over time.

social robotics
: Robot design philosophy that seeks to build robots that act according to human social conventions.

animal-machine interaction
: Interaction design that includes other species in the rules of exchanges with computational machines.

Short bios

A graduate of Carnegie Mellon's Robotics Institute, Marc Böhlen worked as technical research staff at IBM Research Labs before joining the University at Buffalo. He is Assistant Professor in the Department of Media Study where he directs the MediaRobotics Lab. Marc Böhlen offers technology support, the kind of support technology really needs today: www.realtechsupport.org

JT Rinker is a graduate assistant in the Department of Media Study's MediaRobotics Lab and a PhD candidate in Music Composition in the Department of Music at the University at Buffalo.

Photo credits

All Photographs, diagrams and plots by the author