Volkmar Klien
www.volkmarklien.com
info@volkmarklien.com

Book chapter in
'On Listening', edited by Angus Carlyle & Cathy Lane
RGAP (Research Group for Artists Publications)
ISBN-13: 978-0956902443

The artist and the listening machine

In recent decades substantial work focusing on listening technologies emerged from several very active fields of research. As ready to use tools stemming from this basic research are slowly becoming available to sonic artists, it is important for artists and researchers alike to continuously reflect on the conceptualisations of human listening underlying these machine listening approaches and their potential implications in sonic art practise.

In the scientific community there doesn't exist a single and uniform area of research referred to as 'machine listening', attempting to replicate human listening as such. The reasons for this are not primarily to be found in the lack of adequate technological solutions or machines fast enough to tackle the task. It simply cannot be taken as understood what human listening is, how it works, and how clearly it can be distinguished, not only from other modes of human perception and cognition but also in terms of the individual listeners in their physical environment and cultural contexts. There is no single routine (or finite set of routines for that matter) of aural perception that could be automated by being modelled in code in order to replicate its functionality.

Scientists specialising in listening technology related research come from a multitude of backgrounds ranging from psychology, acoustics, musicology and machine learning. They work on providing technological solutions to specific problems related to the realm of human listening, such as the recognition of spoken words, specific songs or sounds[1]. Even though any attempt to pull into dedicated paragraphs what are sizeable and active fields of research promises to over simplify what is mentioned, while omitting numerous other areas of significance, the following will outline three fields, namely speech recognition, music information retrieval (MIR) and computational auditory scene analysis (CASA)[2].

[1] For listening technologies it is normally of no importance if the solutions provided are in any way similar to what a human does when listening. They are judged by their success, not by their similarity to the human auditory system.

[2] A more encompassing overview, accessible also to non-experts, is to be found in 'Sound to Sense, Sense to Sound. A State of the Art in Sound and Music Computing', Polotti and Rocchesso (eds.), 2008, Logos Verl., Berlin

Speech recognition is the listening technology with the longest pedigree. It concerns itself with the transcription of spoken word into text and has come a long way since the presentation of the first tool for the recognition of spoken digits in 1952 [3]. Today commercial packages for 'voice recognition' are readily available and speech recognition is in widespread use in - amongst others - areas of business, telephony and surveillance applications. Yet it is important to remember that converting spoken word to text is not a good analogue for what we normally mean by listening to somebody speak, where we would usually expect a certain understanding of what is actually being said - something that is not in itself part of speech recognition.

MIR is the interdisciplinary science of retrieving information from music in its various representations, such as scores or sound files. MIR researches and provides technologies for searching musical objects, or parts thereof via queries framed in musical terms. One example of this would be a search through a database of songs by humming the hook-line of the specific song looked for. MIR evolves against the backdrop of information society's needs for tools to manage its databases of sound and music, the size of which simply renders unaided 'manual' search unfeasible. Many real world applications have stemmed from MIR's research and are in widespread use, for example in online music recommendation systems[4], in signal-based play list generation systems[5] and music recognition systems[6]. To be able to successfully deliver the answers to user queries MIR does not restrict itself to the processing of data traditionally regarded as 'musical', such as scores and recordings, but instead includes data on how music is used by listeners. It does this by factoring in meta-data, such as that harvested from user tagging or tracking of user behaviour.

CASA is the study of auditory scene analysis by computational means[7]. In essence, CASA systems are "machine listening" systems that aim to separate mixtures of sound sources in the same way that human listeners do. In his seminal book 'Auditory Scene Analysis'[8] Bregman describes the complexity of its task in the following metaphor: "Imagine

[3] Davies , K.H., Biddulph, R. and Balashek, S. (1952) Automatic Speech Recognition of Spoken Digits, J. Acoust. Soc. Am. 24(6) pp.637 - 642
[4] Examples for online music recommendation systems are: Last.fm (www.last.fm), Spotify (www.spotify.com)
[5] The first commercially available signal based playlist generation algorithm was MOTS, which was developed in a co-operation between Bang & Olufsen (www.bang-olufsen.com) and the Austrian Research Institute for Artificial Intelligence (www.ofai.at).
[6] Examples for music recognition systems are: Midomi (www.midomi.com), Soundhound (www.soundhound.com), Shazam (www.shazam.com)
[7] Wang, D. L. and Brown, G. J. (Eds.) (2006) Computational auditory scene analysis: Principles, algorithms and applications. IEEE Press/Wiley-Interscience
[8] Albert S. Bregman, Auditory Scene Analysis, The Perceptual Organization of Sound, MIT Press, 1990, p.5f

that you are on the edge of a lake and a friend challenges you to play a game. The game is: Your friend digs two narrow channels up from the side of the lake. Each is a few feet long and a few inches wide and they are spaced a few feet apart. Halfway up each one, your friend stretches a handkerchief and fastens it to the sides of the channel. As waves reach the side of the lake they travel up the channels and cause the two handkerchiefs to go into motion. You are allowed to look only at the handkerchiefs and from their motions to answer a series of questions: How many boats are on the lake and what are they? Which is the most powerful one? Which one is closer? Is the wind blowing? Has any large object been dropped suddenly in the lake?" Regardless of the problematic underlying concept whereby human perception is interpreted as the building of internal representations of the outside world with the help of sense-data accessible via strictly separated modes of perception, Bregman's scenario (the lake representing the air surrounding us, the handkerchiefs our ear drums) successfully manages to depict the magnitude of the problems in digital signal processing associated with machine listening.

For sonic artists listening technologies in their various guises can provide useful tools. While many of the technologies becoming available today are still in experimental phases it has become clear already that the ability to produce different levels of content awareness in machines will have distinct repercussions not only in the field of the production of sound and music, but also in re-defining the roles these play in everyday life.

By applying these technologies as well as extending their use far beyond their standard application in market driven software solutions, sonic artists can open up new horizons in the production as well as conceptualisation of their work. Sonic installations for example can now be designed to be (to a certain extent) aware of the soundscape they evolve in, enabling them to respond and adapt in real-time. New digital musical instruments may be created allowing users to re-define their individual instrument by simply 'feeding' it sound files containing examples of the sonic qualities required. The list of possible applications is seemingly endless, stretching from instruments reacting automatically to other players to semi- or fully-automatic improvisers and generative music engines capable of mining databases of millions of songs for 'new inspiration'.

On the recipients' side these future technologies will strongly influence listening habits and modes of reception. Tools allowing the user to re-create and re-mix sound recordings or their own live sounding environment, on a perceptually informed level (e.g. 'a car driving by') rather than on a purely technical level (e.g. 'filtering the ~5000hz band'), will help blur the lines between the traditional roles of producer and recipient, shifting the focus further from

static works of art to process-based approaches such as the dynamic re-shaping of real-world soundscapes, extended listening practises and interventions in aural perception.

But it would be naive for sonic artists to uncritically celebrate the new tools simply as means to further improve their bachelor machines. (And yet, were they unable to listen, could they ever truly obey?)

Regardless of how desirable some potential applications of listening technologies in sonic art contexts may be, artists cannot simply rely on scientists to explain and hence also define what listening is. It is important for artists to join the debate and consider the fundamental methodological differences between science and art and the possible implications any unreflected use of technology might have through inadvertently 'importing' implicit conceptualisations of listening into artistic practise. This is ever more important as any formalisation of human listening processes is always based on fundamental ideas about human perception and hence on human nature as such.

From the viewpoint of science, listening might at first glance be (and often enough still is) interpreted as dealing with the perception and interpretation of sound pressure waves. But listening is rather one form of human interaction with the environment as a whole. It is this quality of human listening that enables that enables sonic art to be such a powerful medium for the exploration of human perception and interaction with our world. Especially in artistic contexts (and listening in any context can be this art), listening needs to be understood as a creative act and as such as an open-ended affair. Human listening is not simply the recognition of defined patterns in data retrieved from the outside world via our ear canals. It is not an acoustic land survey, but much rather an integration of the heard into individual horizons at positions not necessarily pre-determined.

As contemporary sound arts do not exhaust themselves in providing audible objects in time by making use of traditional musical instruments or loudspeakers, artists strive to create situations and experiences allowing listeners, these fellow artists, not only to explore their world aurally in novel ways but also to aim at suspending engrained listening habits to expand and sharpen individual listening practise. It is with this focus on listening in mind that machine listening research can be seen as a provider of potentially powerful tools to sonic art practise, but additionally as a neighbouring, if methodologically rather distinct, discipline in the study of human listening in its multitude of contexts.