***Ave Verum Pentium*: singing, recording, archiving and analysing within the digital domain**

Evangelos Himonides

In this chapter, I consider the role of technology in recording, processing and archiving the singing voice. The novel conceptual 'compass' for the present discussion is that the recording chain (i.e. the set of individual technologies involved between the singer's lips and the listener's ears, during recording and playback) is often presented in the literature as deterministic and free of context. In reality, practising singers and recording performers, music producers, teaching practitioners, educators, researchers, and/or people that play multiple roles, often have different needs both in terms of the technological solutions that they require but also in terms of the level of scientific understanding required for effective practice. With this chapter we do not attempt to trivialise the complex worlds of acoustics, psychoacoustics, mathematics, engineering, computer science, performance, pedagogy and production, nor offer a *passe-partout* that unlocks all possible practices and creative foci. It is hoped that this work offers a bridge between the sciences, arts and the humanities, thus allowing readers from different backgrounds to form a somewhat broader understanding about the wonderfully diverse world of the recorded voice and offer insights into (and share challenges about) proximate worlds and practices. It is emphasized that outside the highly specialised worlds of research and scholarship in acoustics, electronics engineering and physics, it is perfectly achievable to perform successful recordings given that we maintain a systematic approach to our methodologies and praxes.

What is sound?

Sound! In almost every book on acoustics and psychoacoustics, there is an opening section that refers to 'sound' being around us, 'with' us, constantly. Even before we are born, during the last trimester of pregnancy, our auditory system is functioning (Welch 2005a; Malloch and Trevarthen 2010). Research also suggests that due to our connection with our mothers pre-birth and the ability to hear sounds coupled with their emotional 'potential' through the bloodstream, during the last trimester in the womb, we enter this world 'pre-programmed' to like and dislike particular sounds, certain melodies and to recognise familiar timbres. Our understanding of the outside world is strongly connected to how things 'feel', 'look' and 'sound'.

But what is 'sound'? What are its properties? How do we perceive 'sound'? How do we visualise 'sound'? Are there common misunderstandings regarding 'sound' and its representation?

According to Everest (2001), depending on the perspective used (what Everest calls ‘the approach’ be it physical or psychophysical), sound can be defined as a wave motion in air or other elastic media (stimulus) or as that excitation of the hearing mechanism that carries out a preliminary analysis of the incoming sound for the perception of sound (sensation). He further explains that “the type of problem dictates the approach to sound. If the interest is in the disturbance in air created by a loudspeaker, it is a problem in physics. If the interest is how it sounds to a person near the loudspeaker, psychophysical methods must be used” (p. 1). Similarly, Howard and Angus (2016) suggest that,

> at a physical level sound is simply a mechanical disturbance of the medium, which may be air, or a solid, liquid or other gas. However, such a simplistic description is not very useful as it provides no information about the way the disturbance travels, or any of its characteristics other than the requirement for a medium in order for it to propagate (p. 2).

A useful metaphor employed within numerous textbooks on physics and acoustics, that can help people understand the propagation of sound, is that of the ‘slinky'. The slinky is a very good way to aid understanding of and to demonstrate wave motion. This visual (and haptic) metaphor can exemplify the two major categories of waves; the 'longitudinal' waves (sound waves are longitudinal waves) and the 'transverse' waves (often found in the vibrations of strings or membranes). The difference between the two can perhaps be clarified using (or imagining using) a slinky in two different ways (see also: The Physics Classroom, n.d.): first, if we rested the slinky on a table surface and held each end with our hands and —while keeping one of our two hands steady— started moving the other hand back and forth, we would achieve a motion that is similar to a longitudinal wave. Alternatively, if instead of moving one of the ends back and forth (i.e. to the axis defined by the length of the slinky), we tried to make small perpendicular movements (i.e. perpendicular to the length of the

slinky) we would be seeing something that is close to what some might know as a 'sinusoidal curve', as long as we tried to keep our movements uniform. This is a good representation of a transverse wave. As mentioned earlier, sound waves are longitudinal waves (i.e. what was achieved with the first slinky experiment).

Sound is readily conducted in gases, liquids, and solids such as air, water, steel, concrete, etc., which are all elastic media. Perhaps one remembers as a child hearing two sounds of a rock striking a railroad in the distance, one sound coming through the air and one through the rail. As Everest (2001) explains: "The sound through the rail arrives first because the speed of sound in the dense steel is greater than that of tenuous air. Sound has been detected after it has travelled thousands of miles through the ocean. Without a medium, sound cannot be propagated. In the laboratory, an electric buzzer is suspended in a heavy glass bell jar. As the button is pushed, the sound of the buzzer is readily heard through the glass. As the air is pumped out of the bell jar, the sound becomes fainter and fainter until it is no longer audible. The sound-conducting medium, air, has been removed between the source and the ear. Because air is such a common agent for the conduction of sound, it is easy to forget that other gases as well as solids and liquids are also conductors of sound." (p. 5).

Capturing sound

Background

In order to form a better understanding about current recording techniques, it might be useful to look at the history of recording. At the time of publication (i.e. 2018), recording sound for later playback had been available to humanity for 140 years. Given a plethora of evidence (e.g. Mithen 2006) or hypotheses (Welch 2005b; Himonides 2012) that humans are musical by design and have been singing and making music since the very beginning of their phylogenetic journey, sound recording can be viewed as an affordance that is practically contemporary. In 1877, Thomas Alva Edison applied to the United States Patent and Trademark Office in order to register his invention that could record sound. On February 19 of the following year, Edison's invention received official approval as patented technology, with patent nr 200521 and the official title 'The Phonograph or Speaking Machine'. Edison's

technology was rather crude, and the quality of playback was quite poor and ephemeral by today's standards, due to the choice of tin-foil as the material on which the vibrations of a needle/stylus caused indentations. A decade later, and thanks to the development work by Alexander Graham Bell and Charles Tainter, a significantly better material, wax, spread on the rotating cylinder of the phonograph, allowed for much better recording and reproduction qualities, as well as longer 'shelf life'. Following on from Edison's original invention, its advancement by Bell and Tainter, and further polish by Edison, a different technology appeared that changed the face of music and sound forever. This was the 'gramophone', invented by Emil Berliner (1887). It is remarkable that the vinyl record, a direct offspring of the gramophone, is still used today, and surprisingly seeing an impressive increase in global sales (O'Connor, 2018). It is a celebration of the importance of the human voice that both technologies that marked the beginning of the era of sound recording used *phono-* and *-phone* in their names, where the ancient Greek word φωνή (i.e. phōnḗ) means 'voice'. Therefore, at birth, sound recording was seen by its forebearers as *voice recording*.

Analog domain

In the natural world, as briefly described earlier, sound exists within what we call 'the acoustic domain', as a strictly mechanical phenomenon. It is important to understand that no different types of sound exist (e.g. analog sound and/or digital sound) as is often misunderstood. Different 'representations' of sound though do exist. These 'representations' allow us to capture, store, edit and replicate performances at different levels of accuracy, at different costs and logistical complexity, with varying levels of reproduction fidelity, at varying levels of perceived warmth and perceived quality, with different levels of complexity with editing and manipulation, and with varying technical and logistical requirements for storage, archival and communication.

As a first step onward from the acoustic domain, we have the 'analog domain' (nb: the American spelling is almost exclusively used globally in this context). Within the analog domain, sound vibrations are converted into varying electrical signals, which are then usually stored onto a magnetic medium, like tape (e.g. reel to reel tape, 8-track tapes, standard cassette tapes). Interestingly, magnetic media are also being used for the storage of digital information (see next section). This introduces multiple advantages, as these

electrical signals can be created using other than traditional voice and/or instrument recordings, like for example with the use of analog synthesizers (i.e. where we do not convert actual sound into signals, but where we create signals artificially in order to convert them later into sound). At the stage of playback, whatever domain we have been working in, we *always* need to move onto the acoustic domain; otherwise we will not hear an audible result. Within the analog domain, in order to play representations of sound back we need to convert electrical signals back into vibrations. A typical means for achieving this is the ubiquitous 'loud-speaker' (or speaker). The speaker is a typical example of what is called a 'transducer', where one form of energy (electrical) is converted into another one (mechanical — the vibration of the speaker membrane, which results in the production of sound).

Microphones

A very important technology within the analog domain is the microphone. The microphone essentially performs exactly the opposite job of that of the speaker. It converts mechanical energy (i.e. sound — vibrations) into electrical energy (i.e. a fluctuating electrical signal).

Paul White's introduction within his short book 'Basic Microphones' (2003) is a very helpful starting point and essential reading for the reader who would like to discover more detail about how microphones work, and how they are used in various recording contexts and situations. He explains: "no matter how sophisticated computers or synthesizers become, the recording of 'real' sound always starts with a microphone. The problem is that, unlike the human ear, there is no single microphone that is ideal for all jobs - microphones come in many types and sizes, and all are designed to handle a specific range of tasks. The problem is in deciding what microphone to choose for a particular application. Having selected an appropriate microphone, there is still the question of how best to position it relative to the sound source in order to capture the desired sound" (p. 11).

Within this overview chapter, we shall not go through the different technologies and designs of microphones in detail. We will simply mention the two major classification factors and briefly explain their basic differences, as they are quite important within the voice recording studio.

One important classification factor for microphones is 'directionality'. Not all microphones pick up sound in the same way, and the type you choose will depend on the task at hand. Some pick up sound efficiently regardless of the direction from which the sound is coming — in other words you don't have to point the microphone directly at the sound source because it can 'hear' equally well in all directions. Some microphones may be designed to capture mainly to sounds from a single direction while others may pick up sounds from the both front and the rear but not from the sides. These basic directional characteristics are known as:

- omnidirectional (all directions)
- cardioid (unidirectional - literally 'heart shaped')
- figure-of-eight (mics which pick up from both front and rear but not from the sides)

(White op cit, p. 16).

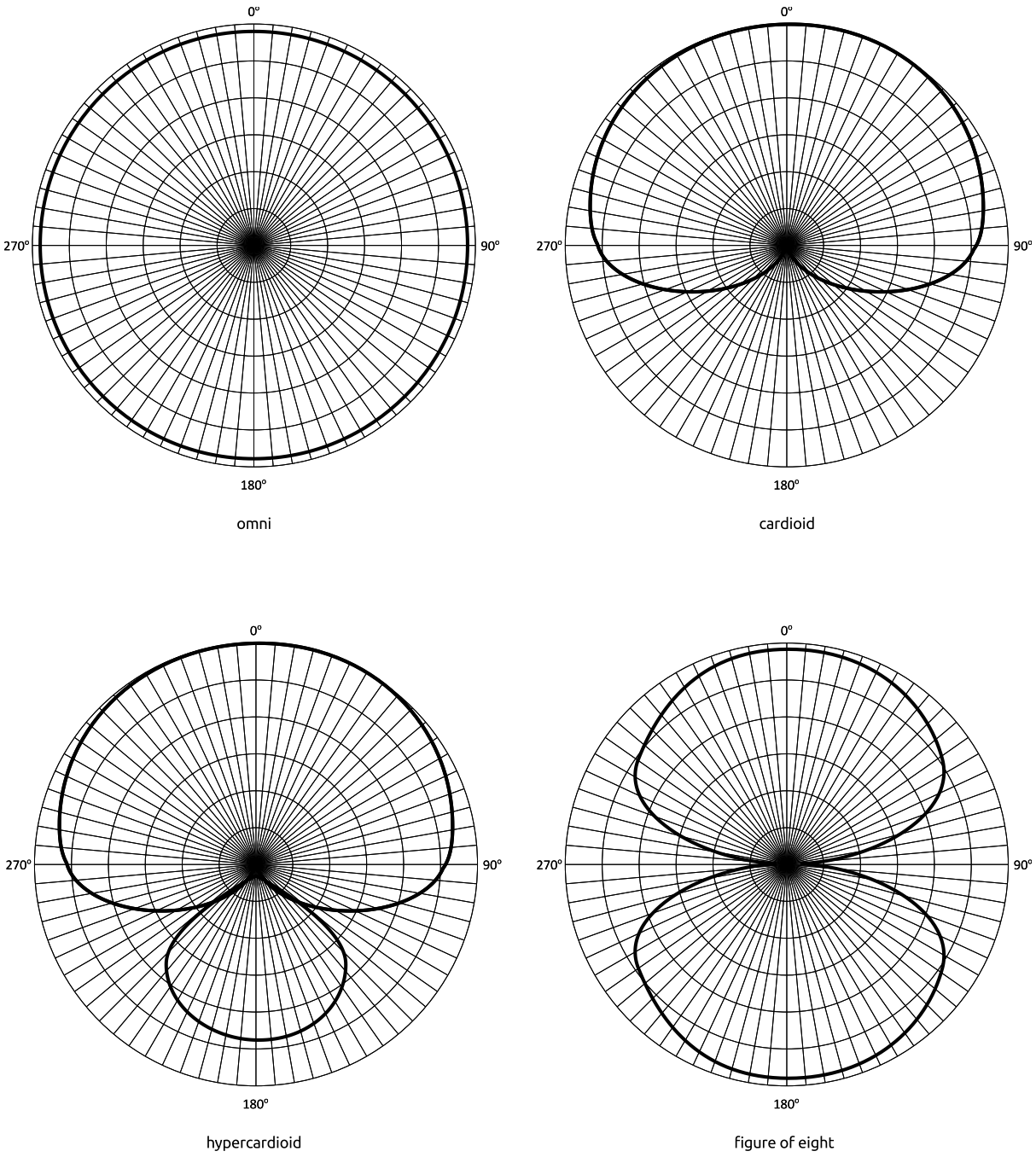


figure 1: common microphone polar patterns

Secondly, based on the construction of the microphone itself (i.e. its 'topology') we can have:

- dynamic microphones
- ribbon microphones
- capacitor (or condenser) microphones

The differences between the three are related to the slightly different topologies employed in order to convert sound to electricity.

Different designs result in not all microphones behaving similarly, and not all microphones showing similar qualities in their operation, the accuracy of the signal that they generate (see above, as the *representation* of the recorded sound). Microphone design is a very complex science, but also a praxial battlefield for very passionate exchanges based on empirical convictions that recordists, engineers, singers, researchers and practitioners possess. Remarkably, microphone technology is also not a plateau where we have seen much progress and innovation in the past century. Some of the most expensive, valued and cherished studio microphones to date are microphones that were manufactured before WWII with countless contemporaries trying to imitate their design, at varying levels of success (see for example Neumann/Telefunken and Gefell microphones).

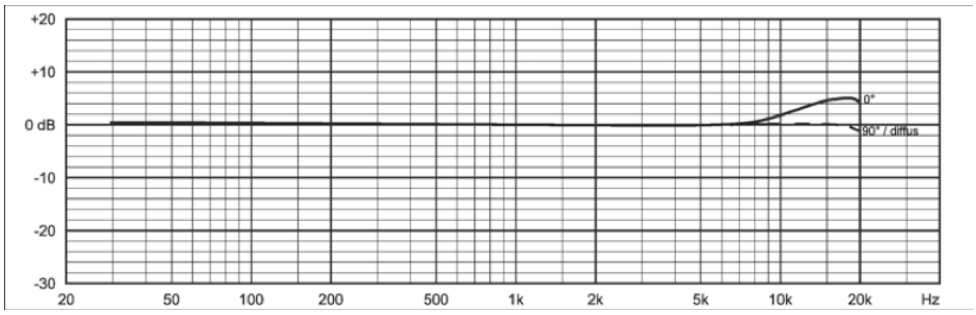


figure 2: the frequency response of the Beyerdynamic MM 1 professional grade measurement microphone. Here we can see the remarkably flat frequency response throughout the devices entire range (20Hz – 20000Hz).

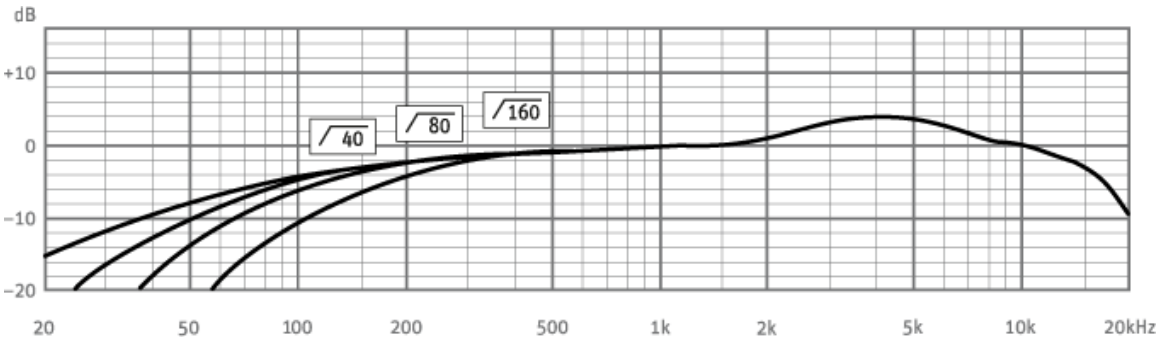


figure 3: the frequency response of the neumann M149 tube microphone at a figure-of-eight configuration and four different low frequency 'roll-off' settings. Here we can observe

how the microphone is tuned to boost frequencies within the 2-6KHz band, which is something that is perceived to add 'warmth' and 'presence' to vocal recordings.

As with other contexts where we attempt to 'capture' a phenomenon, like photography for example, there is no ultimate or definitive tool that we can use in order to achieve optimal results. In reality, we always need to perform certain compromises in one area in order to gain better results in another; we need to consider the cost of our decisions; we need to adjust our 'gains' in order to achieve optimum 'yield' within our set of logistical constraints and affordances. Since we mentioned 'photography' above, it would be interesting to perhaps remind ourselves that the definitive lens for a camera does not exist. A lens that is extremely fast, sharp, and responsive at larger distances is a lens that would require extremely expensive optics, but also a lens that would need to be terribly heavy and large in size compared to a standard fixed 50mm lens. On the other hand, a lens that would work brilliantly for long distances could never outperform a lens that was designed for macro photography (i.e. close up photography) or a lens that would produce acceptable results for most photography tasks (an 'all rounder') and a range of foci.

In recording, a microphone that has been designed for ultimate 'accuracy' in capturing sound (in this context, meaning a microphone that has as flat a frequency response as possible) is not necessarily a device that would be appropriate for recording real singing performances. Singers and producers only care about perceived 'warmth', 'presence', 'punch', 'distortion', 'colour', and 'character'. All these properties are usually opponents to 'accuracy'. One would therefore not find a lab 'measurement microphone' in a commercial recording studio used for recording artists. This is why we would rarely see an omni-directional microphone being used for recording studio vocals, as opposed to cardioid or hyper-cardioid designs, whilst it would be inappropriate to use anything but an omni-directional microphone for acoustic analyses, measurement and research.

Similar 'compromises' (or 'tactical decisions') have to be performed with the choice of microphone capsule or diaphragm size. Once again, no definitive answers can be offered to the question "which microphone should I use?" There are numerous factors at play with the different capsule designs and sizes, that result in varying ability to handle large sound

pressure levels, but also in varying self-noise levels, and in varying success levels for perceived proximity effect (i.e. the perceptual assessment of how close to the capturing device the sound source had been). For example, tiny diaphragms can handle much higher sound pressure levels, yet at the same time they would suffer much higher self-noise levels.

To date, one of the most comprehensive scholarly works that present the complexity of microphone technology within the voice research context is that of Jan Švec and Svante Granqvist (2010) for the American Speech-Language-Hearing Association. Therein, the authors offer evidence about how different design aspects of microphones impact systematic acquisition of recorded data in great detail.

In the present context, it is useful to consider the following, depending on the recording task:

educators / practitioners: when recording singers for reference, and/or in order to gauge development longitudinally, the most important thing to safeguard is a replicable and systematic recording, as it is perceived change that matters, keeping the underlying technologies for recording it constant. Nowadays, one can achieve something like this with very accessible and affordable technologies. Things to consider when trying to keep a constant are: selection of microphone, microphone placement, recording venue / room (i.e. room acoustics), positioning of the singer in relation to the microphone but also within the room, pre-amplifier gain settings (see hereafter: microphone pre-amplifiers), and remaining recording chain (see hereafter: the digital domain). Following the previously offered argument that very little has changed in microphone design within the past century, it is interesting that what this author perceives as one of the most exciting recent developments in microphone designs since the introduction of the microphone is not really related to the sound capturing technology. It is the introduction, by British manufacturer Aston, of a laser pointing device built onto the body of their 'Starlight' model microphones, thus enabling the user to perform very accurate positioning and replicate placement from one recording to another.

researchers / scientists: within systematically researched contexts, vocal recordings need to be calibrated, systematically designed, set up and conducted, but also using high quality and reference technologies that enable the researchers to 'capture' sound (i.e. the acoustical phenomenon) as accurately as possible within the analog and digital (see below) domains. In this respect, researchers should solely be using high quality reference microphones, with omni-directional response patterns, extremely flat response curves, very low self-noise, and placed systematically at the appropriate distance from the singers' lips and within carefully controlled singer placement and room acoustics. Beyond the microphone, the remainder of the 'recording chain' would also need to be carefully controlled. This means that we need to understand the role of microphone pre-amplification, as well as how the analog signal becomes converted into digital 'information' (nowadays, almost exclusively!)

artists / producers: it is oft celebrated by successful producers and sound engineers that "if it sounds right, it is right". Experience and studio practice suggest that there really is no point in immersing ourselves into science in order to achieve a successful recording. Any microphone is a potential candidate in this context, especially within popular music genres. Accuracy, transparency, fidelity and/or clarity are often not seen as important, when distorted, coloured, lo-fi or branded sound recordings are sought under a specific aesthetic paradigm. This can further become exaggerated within editing and post-production. Within this context, the technological continuum is so vast that one can witness multi-platinum vocals that had been recorded using the microphone of a discarded telephone handset, but also recordings that were performed using vintage Neumann U47 large diaphragm condenser microphones (often selling for close to 10,000 American dollars).

Microphone preamplifiers (also known as 'MIC PREs')

Microphones put out small voltages; desks and outboard gear work at much higher voltages. A pre-amplifier is therefore needed between a microphone and whatever 'capturing device' to turn the millivolts at the mic output (mic level) into volts for processing (line level). This simple task is absolutely central to the recording process. Any signal that 'leaves' the microphone will be affected by the design and quality of the mic-pre that receives it, and any noise, distortion, or inaccuracy introduced to the sound at this point will become a permanent part of the signal (or 'the information' within the digital domain past conversion

to digital data). Once again, the neighbouring paradigm of photography has been used by famous producers such as English producer and engineer John Leckie, in order to demonstrate the importance of the microphone preamplifier within the recording chain. Leckie often appeared to compare the microphone preamplifier with camera lenses, highlighting how important their quality is in determining the quality of the final captured photograph. Within professional recording circles, the importance of microphone preamplifiers is often seen as greater than that of the microphone: High quality mic pre's affect the performance of microphones very directly, and an ordinary mic through a top quality mic pre will sound better than a good mic through a poor mic pre [source: http://www.proaudioeurope.com].

It is worth noting that microphone preamplifiers are not always present as dedicated, stand-alone devices within the recording chain. We can often see microphone preamplifiers built into mixing desks, computer recording interfaces, solid state recorders, live or studio vocal performance units and/or pedals, dedicated recording 'strips', mobile phones, and even concealed within devices that appear to be traditional microphones (e.g. contemporary USB microphones). The latter will also feature built-in analog to digital converters (something that we shall cover hereafter).

In light of this, it is important to consider the different needs for the three general recording 'umbrellas' that we identified earlier:

<u>educators / practitioners</u>: similar to employing an appropriate microphone, the choice of microphone preamplifier is not a complicated exercise, and its use should be focused on being systematic rather than being scientific. As mentioned earlier, it is important to try to filter out possible variables and/or contaminants when we aim to monitor singing development and singing performance practice longitudinally. Therefore, the use of a decent quality microphone preamplifier, even if built into a standalone solid state recording device, is perfectly acceptable, as long as the practitioner has control over its gain settings. Although there is no real need for properly controlled calibration of sound pressure levels for the recording within this context, it is essential that the gain settings are set (i.e. not automatically adjusted by the recorder – known as 'auto-gain'). Additional care needs to be

offered in ensuring that the amplified signal is not overloading the analog to digital converter (we shall clarify this in the following section).

researchers / scientists: a microphone preamplifier within the scientific research context needs to be as close as possible to what many sound engineers call the hypothetical 'straight piece of wire with gain'. This is presented as 'hypothetical' because, once again, practice suggests that no hardware design topology can actually result an amplifier that can achieve this perfectly (i.e. to take a low level signal generated by the microphone, and simply make it 'louder' without any alteration). Simplistically, if we performed comparisons (i.e. spectral analyses) between the un-amplified and the amplified microphone signals using a theoretically 'perfect' microphone preamplifier, we should not be able to see any spectral differences. Practically, this is not achievable. The design of any amplifier will have an impact on the distortion and/or 'coloration' (i.e. the *change*) of the resulting amplified signal. Once again, a compromise is required so that we can use a sensible, but also affordable, technical solution. Additionally, sound pressure level calibration (i.e. SPL calibration) within this context is absolutely essential. This is not only because preamplifiers will affect/brand the resulting signal differently depending on the device's gain settings, but because we will not be able to make any valid assessment about singing energy, energy slope and/or subglottal pressure levels during singing unless we have a systematic reference of what the amplifier 'contributed' to the final signal (or digital representation of it – see below) at the time of analysis. For further details, see Švec and Granqvist (2010).

artists / producers: similar to the microphone paradigm, we once again face a praxis where 'everything goes'. Any type of amplifying technology can be used, and often misused or abused, in order to foster creativity. Producers and engineers have been known to utilise any type of amplification technology in trying to create novel sonic products, from guitar amplifiers, to old vacuum tube (aka valve) radios, to PA systems, to low fidelity amplifiers, all the way up to 'boutique' and significantly expensive topologies that can be found in professional mixing desks (e.g. SSL, API) and dedicated, standalone, pre-amplification units (e.g. Manley, Great River, Millennia et al). Different aesthetic 'schools' exist within the recording and audio production worlds, and these are strongly reliant on different types of microphone pre-amplification designs (e.g. from the 'glassy' and transparent pop diva type

vocals, to the 'edgy' and 'punchy' Nashville country vocals, to the 'brown' and over 'saturated' British pop vocals).

Digital domain

One can be quite confident in claiming that we have all used the term 'digital'; some of us on a daily basis! Nevertheless, it would be useful to remind ourselves what we actually mean when we refer to 'digital technologies' and (somewhat erratically) 'digital audio'.

How many times have you heard the phrase "...we live in a digital age..."? How many times has a sales-person tried to persuade us that a product is bound to be "better" because it's digital? How many times have you heard [or taken place in] discussions about analog -vs- digital, the purity [or warmth, or 'thickness', or 'creamyness', or substance, or colour, or quality, or depth, or richness, or...] of vinyl (the records, not the school of fashion) compared to CDs? Many times I should assume... But what is digital? And, consequently, what is what some people refer to as digital audio?

A digital system is one that uses discrete values rather than continuous values: compare analog. The word comes from the same source as the word digit: the Latin word for finger (counting on the fingers) as these are used for discrete counting[1]. In circuitry, a digital circuit is one in which data-carrying signals are restricted to either of two voltage levels, corresponding to logic 1 or 0 (see among others: wgcu.org). In terms of technology in general, digital describes electronic technology that generates, stores, and processes data in terms of two states: positive and non-positive. These two states are described by the two available symbols of the binary system. Thus, data transmitted or stored with digital technology is expressed as a string of 0's and 1's. Each of these state digits is referred to as a bit (and a string of 8 bits that a computer can address individually as a group is a byte)[2].

No matter how complicated the software running on a computer, ultimately everything is being translated into zeroes and ones. This is how computers work. Machines with digital circuits only operate on this binary logic (1-0, yes–no, positive–negative).

[1] source: wikipedia.org
[2] source: iptv.org

Some inventions prior to the appearance of computers have claimed to be the 'ideas' that led to the conception of the first computer. Based on the same binary logic, industrial sewing machines could be programmed in order to produce different designs and patterns. The sewing heads were controlled by a perforated paper-tape. When the tape that was feeding the head at a given moment had a hole, the head would move down, otherwise it would stand still. Many devices used this hole/not-hole technology, either in a single linear fashion (just one line of holes or gaps) or in a multiple line fashion.

What is very important for us to understand is that digital is nothing but a *representation* of data. In the case of audio and sound, digital audio is a *representation* of sound and *not the sound itself*. According to what we presented within the introductory section sound is a physical [mechanical] phenomenon... there is no such thing as digital sound; if we are able to hear something, then it is definitely an acoustic (i.e. mechanical) phenomenon.

You can now easily understand that —fundamentally— a representation of a phenomenon cannot possibly be better than the actual phenomenon; it can be, though, an extremely accurate representation of the phenomenon, and in many cases, it can be so accurate that the benefits for utilising such a representation can be immense.

- the representation (successful or not) can be replicated faithfully and effortlessly;
- the representation can be distributed through various channels of communication;
- the representation will not change;
- the representation can be easily manipulated, edited and altered deterministically;
- the representation can be easily archived;
- the representation can be easily retrieved.

Sampling

Since we have established that digital is a representation of a phenomenon and not the actual phenomenon, we need to be a little bit more analytical about how we represent a phenomenon. There are various metaphors that people use in order to explain digital and over the years there is one that this author has developed and become particularly

accustomed to using with his students. Imagine that you witness a crime and that you go to the nearest police station in order to report it. Some police stations employ sketch artists who liaise with the witnesses in order to draw images of the criminals. You have to describe the person that you've seen to the artist and you have to do it in a fixed period of time. Obviously, the best thing that could possibly happen would be for you to produce the actual criminal and show them to the artist. But this is not usually possible. Given that you have a fixed amount of time to describe the person, it seems that two issues are of the essence: The amount of information that you will give to the artist each time (i.e. how long your sentences are going to be every time you open your mouth) and the number of times that you will do this (i.e. how many sentences of XXX length per unit of time). In theory, if you possess a photographic memory as well as remarkable linguistic skills, your description could lead to an extremely accurate representation of the criminal... you could go into so much detail that you are describing each pore of their skin! In any case, the more information you provide and the more times you provide this information will produce a better [in terms of a more realistic] result. This leads us to the two most important aspects of sampling (i.e. what we do in order to produce a digital representation of an analog phenomenon):

BIT-DEPTH or WORD-LENGTH which is the amount/size of information that 'we' provide/store each time 'we' describe the phenomenon; and,

SAMPLING RATE or SAMPLING FREQUENCY which is the number of times per second that 'we' provide the above mentioned 'chunks' of information per second.

Binary, bits and bytes

Since computers can only 'perceive' things in a binary fashion (see above), all information that is being 'fed' into them, needs to be translated to binary code (i.e. into zeroes and ones). Understanding how this works requires a very short refresher from our primary school years, as augmented by the introduction of remedial algebra during high school).

Since our early years, we have been educated and 'branded' to understand numbers using the decimal system. The decimal system is nothing more than another convention so that

people could have a common ground for describing, exchanging and utilising information. The base of the decimal system is, of course, the number ten (10). The numbers [digits] that can be used in the decimal system are 0,1,2,3,4,5,6,7,8 and 9. Everything else is a composite using these ten available ingredients.

Take for example the number 157. What does 157 actually mean? Primary-school children learn that 157 means: 7 units + 5 sets of ten + 1 set of a hundred. Later-on in our lives, most of us learn the algebraic interpretation of the same definition which is $157 = (1x10^2) + (5x10^1) + (7x10^0)$

In the binary system, we can only use two numbers [digits], zero (0) and one (1). In order to represent a number in the binary system we follow the same line of thought as presented for the decimal system, with the obvious limitation that we can only work with the only two 'ingredients' (0 and 1) and the powers of our base (the number 2, see table 1). Therefore, the decimal number 157 can be represented as 10011101: $10011101 = (1x2^7) + (0x2^6) +(0x2^5) +(1x2^4) +(1x2^3) +(1x2^2) +(0x2^1) + (1x2^0)$

table 1: the first ten powers of 2

power	symbolism	decimal result
0	2^0	1
1	2^1	2
2	2^2	4
3	2^3	8
4	2^4	16
5	2^5	32
6	2^6	64
7	2^7	128
8	2^8	256
9	2^9	512
10	2^{10}	1024

'Digital audio'

Now that we have a somewhat clearer understanding about how computers process and 'understand' data we can continue with our introduction to audio in the digital domain. As mentioned earlier, in the real world, the sound of our voices for example, is an acoustic phenomenon. During recording, and with the use of microphones, these acoustic phenomena are converted into electrical signals. To process these signals in computers, we need to convert the signals to "digital" form. While an analog signal is continuous in both time and amplitude, a digital signal is discrete in both time and amplitude. To convert a signal from continuous time to discrete time, a process called sampling is used. The value of the signal is measured at certain intervals in time. Each measurement is referred to as a sample[3]. In order to 'convert' an analogue signal into a digital representation of it, we practically perform thousands of amplitude measurements per second and store the amplitude values. This is called sampling.

Please have in mind that the term sampling is also being used in modern music production with reference to the recall of pre-programmed samples (audio snippets) with various triggers (buttons, controllers, keyboards etc.). A modern sampler is a device that stores recorded sounds and is able to manipulate them and reproduce them allowing them to be distributed across a keyboard and played back at various pitches (see for more info: sample-based synthesis). Both sampling-rate and word-length are absolutely *essential* factors concerning the accurate representation of the signal.

Sampling-Rate (or sampling-frequency)

Imagine that you have to describe (or paint) how bright the sky is during a 24-hour day. If you go outside, fix your photo-camera on a tripod, point to the sky and take one photograph at 11pm and just one more after 24 hours (for this example, let's not worry about colour - just brightness - and let's assume that we have a monochrome film) this is how our photographs will look like:

[3] Source: Thomas Zawistowski and Paras Shah, Engineering Computing Center, University of Houston.

figure 4: two snapshots taken 24 hours apart

If we decided to shoot another one at mid-day, then we would probably have:

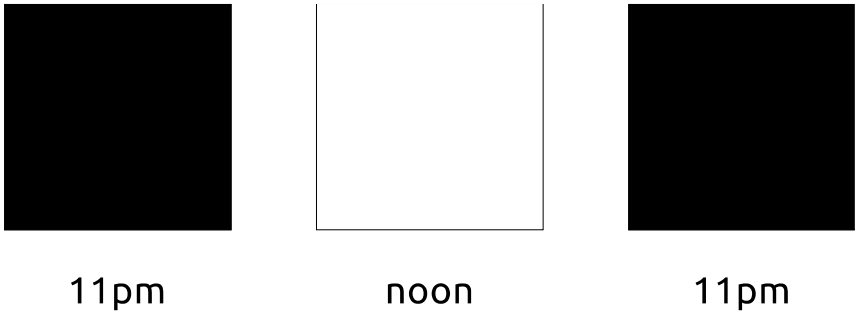


figure 5: three snapshots throughout the 24 hour period spread 12 hours apart

And, of course, the higher the number of photos we take during the 24 hours, the better our understanding about brightness will be:

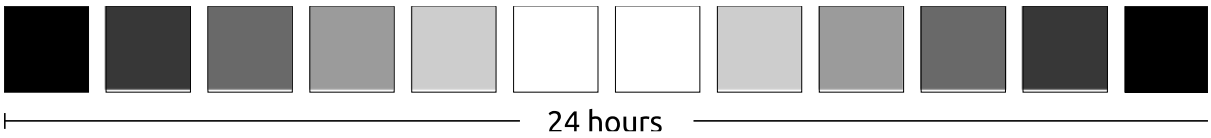


figure 6: a larger number of timed snapshots offers a better understanding of the differences in brightness

A video-camera (which is nothing more than a camera that shoots anything between 30 and 50 photos per second) would produce something similar to:

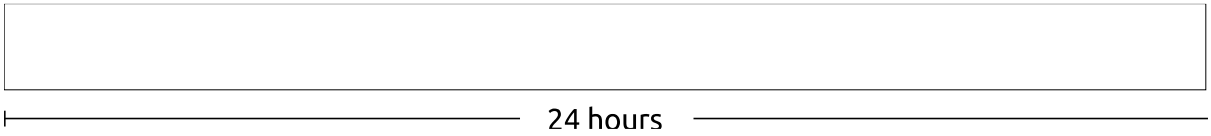


figure 7: a high resolution (i.e. using high sampling frequency) recording of the phenomenon

How many samples are necessary to ensure that we are preserving the information contained in the [audio] signal? If the signal contains high frequency components, we will need to sample at a higher rate to avoid losing information that is in the signal. In general, to preserve the full information in the signal, it is necessary to sample at twice the maximum frequency of the signal. This is known as the Nyquist rate. The Sampling Theorem states that a signal can be exactly reproduced if it is sampled at a frequency F, where F is greater than twice the maximum frequency in the signal [ibid][4].

Some of us perhaps know that 'CD-quality' audio is sampled at 44.1 KHz. This is connected to the fact that our ears are able to 'hear' frequencies from 20Hz to 22,000Hz. According to the Nyquist theorem, in order to represent frequencies up to 22KHz we need to use a sampling frequency greater than twice the frequency in the signal... hence, 44.1KHz.

People with 'golden ears' claim that 44.1 KHz sampling-frequency is simply not high enough. New, high-definition recording and production is using 96KHz (DVD audio standard) and - more extreme - 192KHz sampling frequencies. In theory, the latter is adequate for the exact representation of audio signals up to 95KHz (when human ears cannot possibly hear frequencies above 22KHz). Why go to so much trouble? This is a very complicated field (the field of psychoacoustics); in a nutshell, it is believed (and continually researched) that although it is not possible to 'hear' frequencies above the 22KHz limit, the 'interaction' and 'masking' of higher frequency components with the audible frequencies produces blended results that 'affect' the listener and/or 'trigger' different aesthetic experiences when higher sampling rates are being employed.

Bit-Depth (or word-length)

CD-quality audio (as mentioned above), uses 16bit words for each channel (16bit, 44.1KHz, Stereo). What does this mean? When we are sampling, we store in our machines (44,100 times per second) information (words) that describe the amplitude of the waveform at each given time. Since each word is 16 bits long, this means that we are able to represent a minimum value of 0 and a maximum value of 65,535 when describing the amplitude at a

[4] Further information can be found in a plethora of sources under the keyword " Nyquist Theorem".

given time. What happens when we sample audio using 8bit words? You can see that the possible amplitude values that we can use are significantly less.

Try and 'parallelize' this to the visual-world again... do you remember the very old computer graphics printouts? A pixel could either be blank (white) or black... (couldn't we call this a 1 bit sampling?). This allowed the reproduction of quite crude images where detail was lost due to the limited amount of available colours (or shades of grey). In the following image, we can see the difference between a low resolution and low bit depth image and the same image sampled with a higher resolution and bit depth.

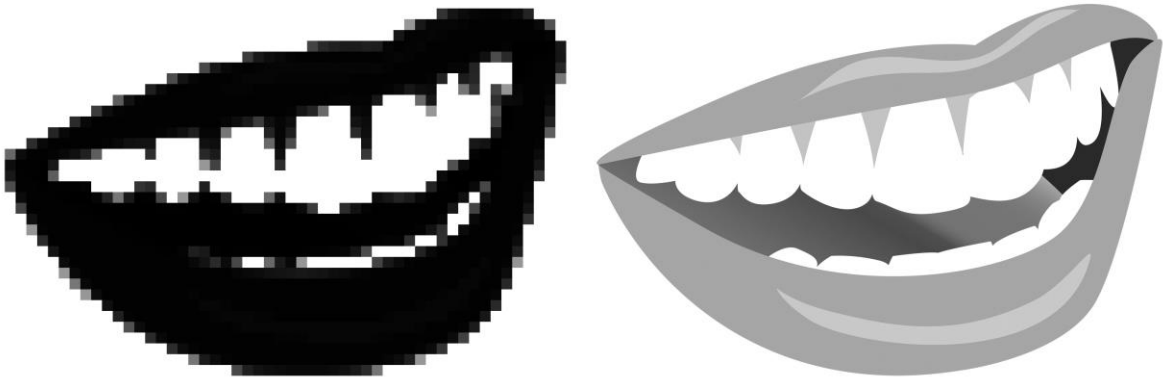
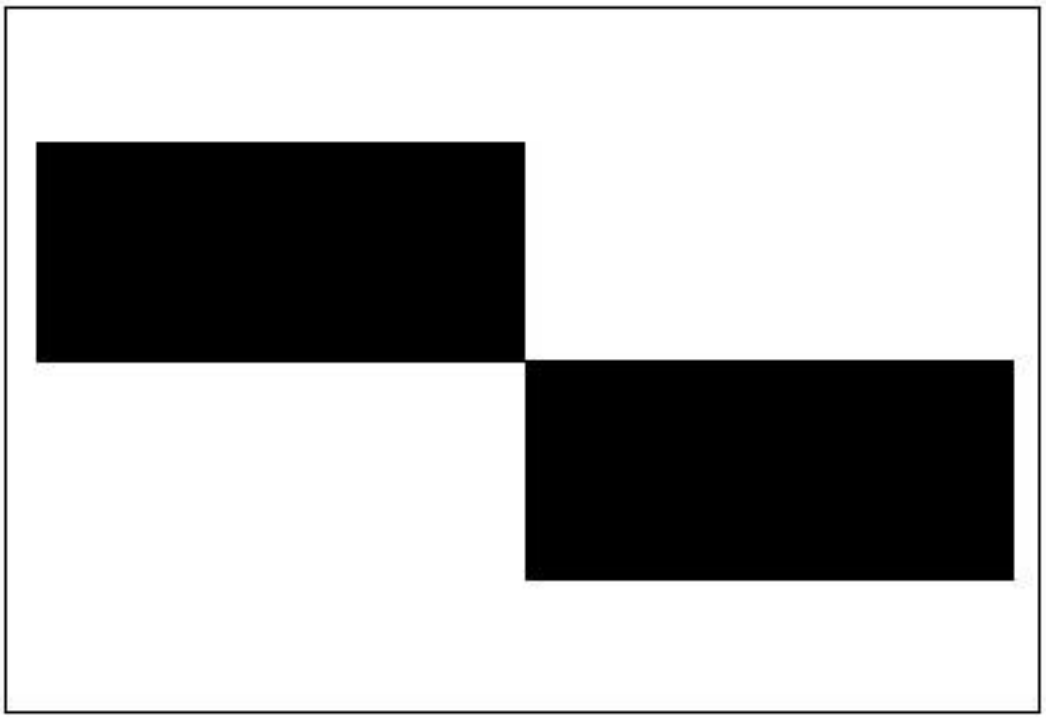


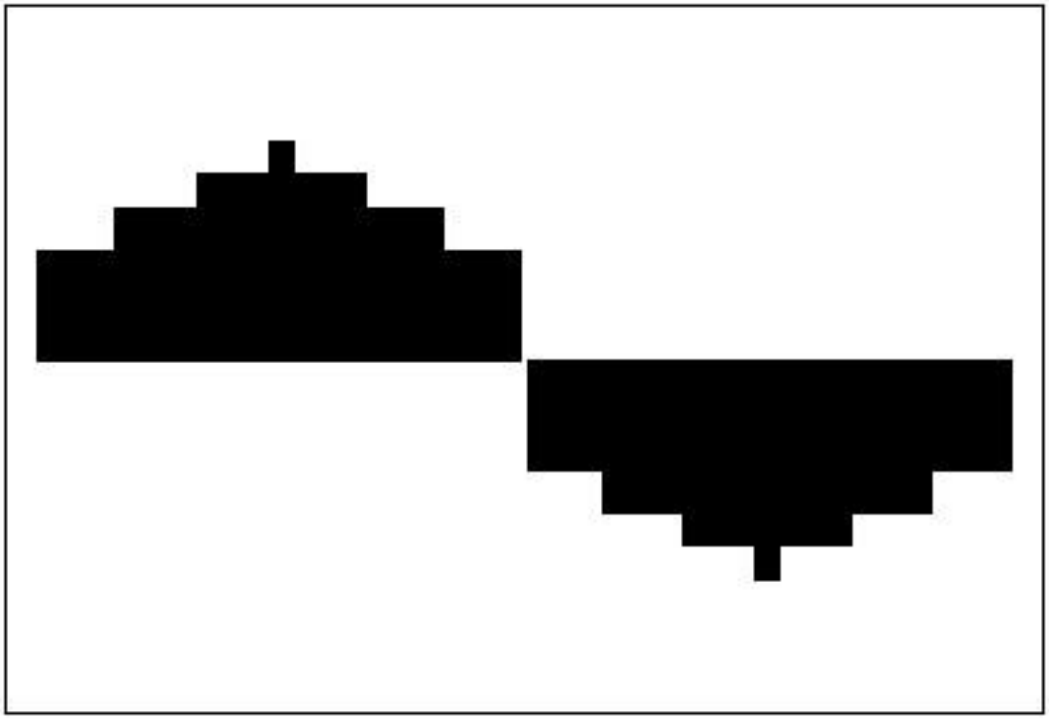
figure 8: comparison between low and high bit depth

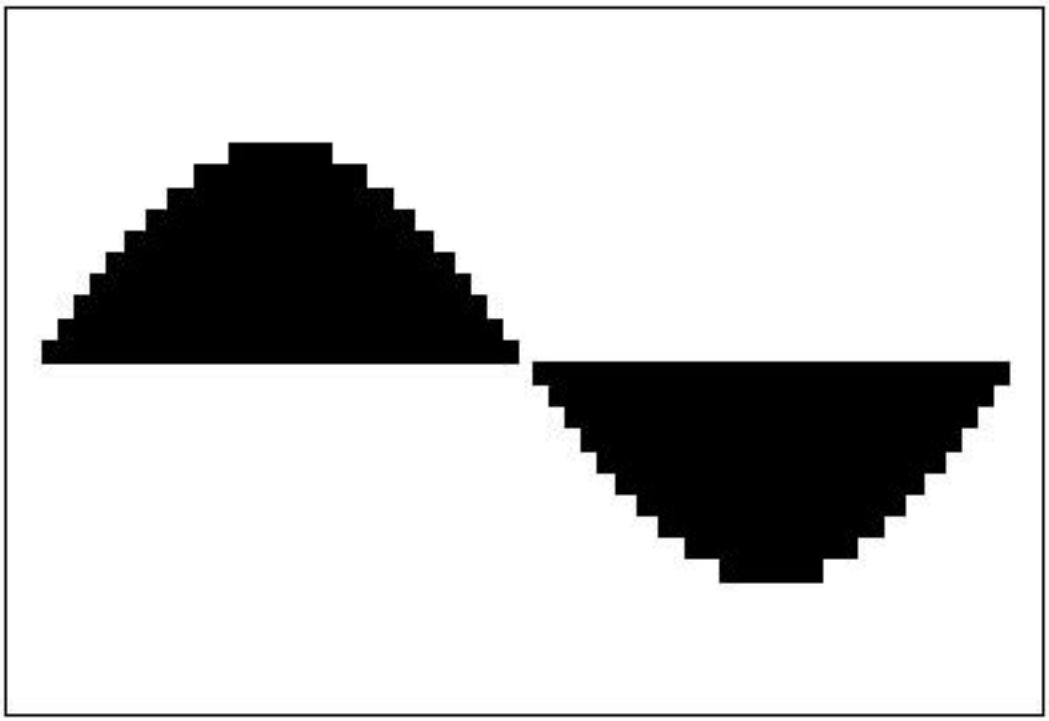
If you open your computer's display properties you will see that your Windows (or Macintosh) system is configured to run at either 24 or 32 bit (!!!) colour quality... [they cannot be bothered any longer to give you the actual number; that is why they say "millions of colours".

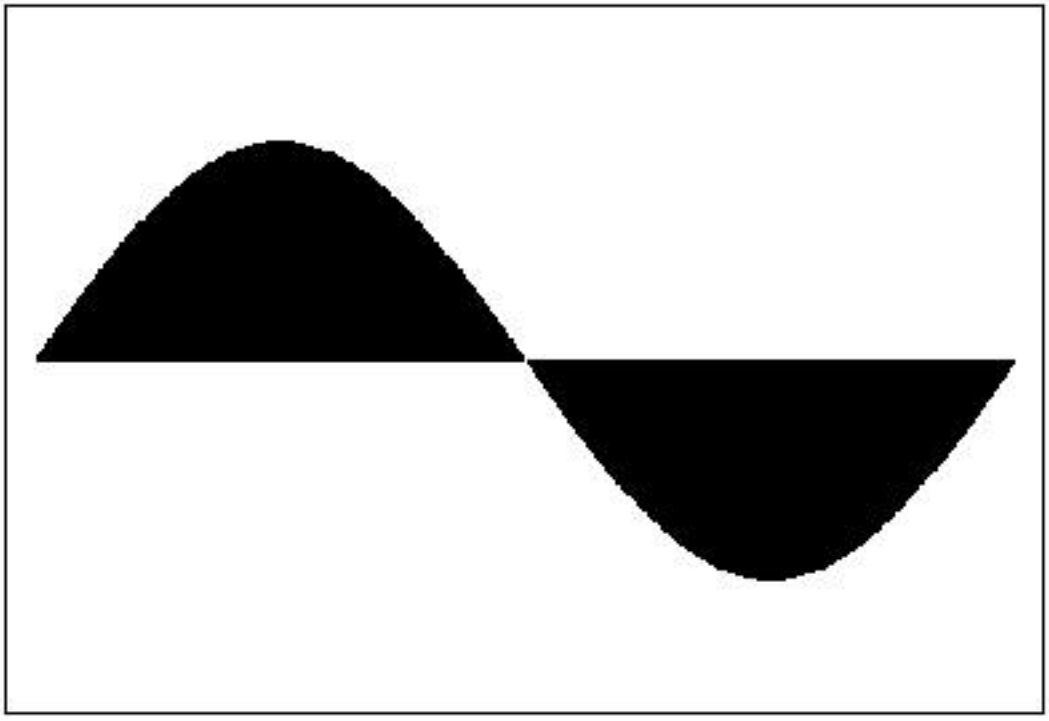
Exactly the same occurs when we are sampling audio:

Depending on the level of detail, we could achieve different levels of accuracy/quality if we tried to sample a sinewave. The graphical representation of such an exercise should look like this:









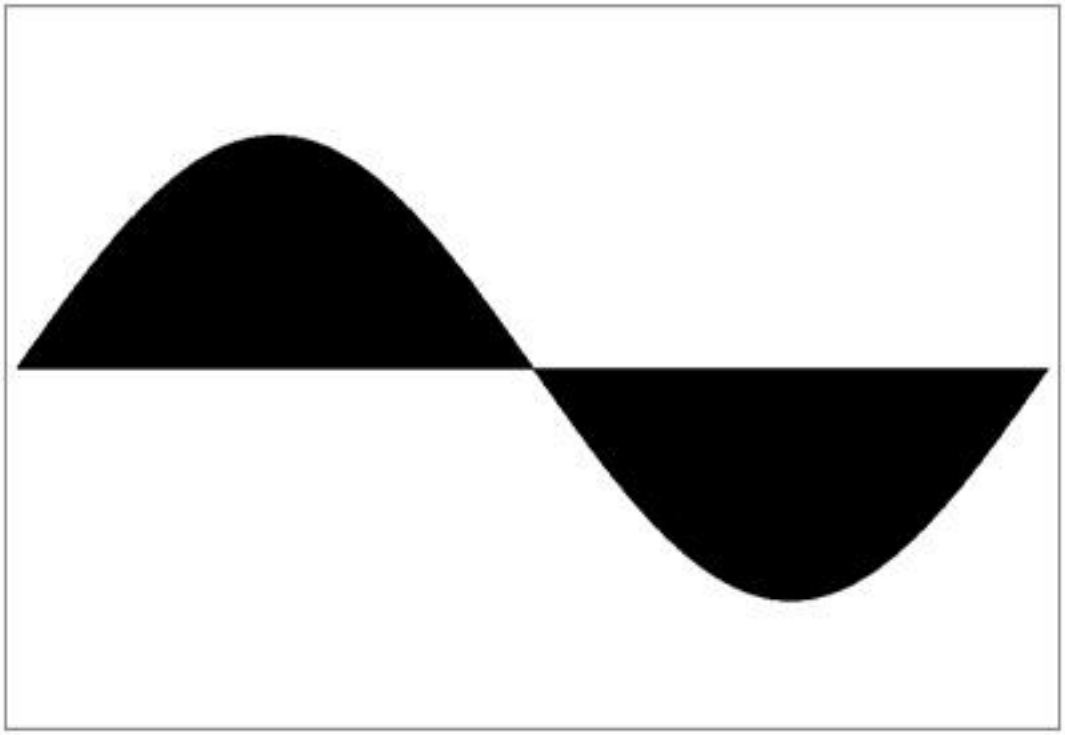


figure 9: an example of sampling the same waveform at different bit depths

You can understand that the quality of the representation is strongly related to the sampling word-length. The bigger the word, the better the sampling...

Moving onto the digital domain

Moving on from our superficial crash course to digital theory and sampling, it would be useful to offer a quick overview of how this occurs in practice. As explained earlier, the acoustic phenomenon (the vibrations, the sound) excite a surface on the microphone (the transducer) which converts mechanical energy (the sound) to electrical energy (electrical signal). This electrical signal is then fed to an amplification device (the microphone pre-amplifier) in order to become stronger. The amplified signal is consequently fed into a device that performs sampling (as described above) at the sampling rate that we have determined (44.1 thousand times per second for CD quality audio) generating chunks of information (words) of an also pre-determined size (bit depth or word length — 16 bits for CD quality audio). The device that performs this type of conversion is called an 'analog to digital converter' (or A/D converter). As with previous components in the 'recording chain', A/D converters can be stand-alone devices, but they can also be 'bundled' inside other devices, even modern microphones (the new generation USB microphones that are

essentially bundles of a microphone, a microphone preamplifier, and an A/D converter all in one). One can find 'hidden' A/D converters in most new technologies, their mobile phones (either used as phones, or as recording devices), solid state voice recorders, video cameras, computer soundcards, built-in audio interfaces for laptop computers, RaspberryPi and Arduino 'shields' and 'hats', wearable electronic components, even cheap plastic toys that allow voice recording (albeit sometimes unpleasantly 'lo-fi').

Past the A/D converter, we solely 'deal' with zeros and ones; as explained above. There is no raw information about sounds or frequencies, solely thousands of measurements of amplitude values stored sequentially. This is an important thing to remember and stresses the need for the following section of 'processing', as some of the things that we do when editing 'digital audio' are related to frequency and timbre. This automatically suggests that one more 'transformation' is somehow necessary in order for us to work with spectral aspects of sound (more accurately, the digital representation of it!) It is not part of this chapter's focus to expand this much further than reporting that one mathematical process that allows us to compute frequency information from amplitude/time sequences is called a Fourier Transformation (from the French mathematician Joseph Fourier). Nowadays, digital data is processed using 'fast' Fourier transformations (FFT). These sit under the thematic umbrella of Digital Signal Processing (DSP). Some readers might have come across music technology equipment reviews where a particular synthesizer's or effect unit's 'DSP engine' was reviewed.

Processing within the digital domain

We have produced a sound (perhaps a beautiful singing performance), captured it with a microphone by converting it to electrical signal, amplified the signal with a microphone preamplifier, fed the amplified signal to a computer sound-card's analogue-to-digital-converter and captured (recorded) the resulting *data* on our computer with the appropriate software (this could be a dedicated 'wave editor' like the free software Audacity, or part of a Digital Audio Workstation (DAW) like Cubase, Logic, Sonar, Nuendo, Reaper, ProTools, Live, Garage Band etc.) What happens past this point depends upon the type of operation one performs onto the stored stream of digital data, exactly as one would manipulate different columns of numbers on a software spreadsheet (like MS Excel). There are countless of

processes that one could apply onto the digitally stored data. Some examples of popular 'effects' (or Fx) available with most software tools are offered below.

Compression. Compressors are sophisticated dynamics processors that sound engineers use, especially for instruments that produce sounds with a great variability in dynamic range (the voice being the most representative). Nowadays, with modern production, practically everything goes through a compressor. In a nutshell, a compressor reduces the dynamic range (making the distance between the loudest and the quietest part smaller). Think of a compressor like being an invisible hand that increases or reduces the gain of a signal in milliseconds (according to our needs). Compressors are also used when we want to remedy recording problems that occur with hard consonants like 'Ts' and 'Ps' and can also work in tandem with equalizers when in some occasions 'Ss' sound harsh (then we call them De-essers).

Equalisation (EQ). Most of you are already familiar with the concept. EQing is an essential part of recording and production. By increasing the energy within specific bands in the spectrum we can facilitate the placement as well as the clarity of the different sounds in the final mix. Some might have heard or interacted with software equalisers and might have also come across different types that somehow resembled (or presented visual metaphors of?) old analog parametric and graphic equalizers.

Reverb. Reverberators are processors (or just algorithms, when used as effects within sound-editing or production software) that change the recording by 'branding it' with the acoustical characteristics of different physical spaces (venues, rooms). This is achieved in number of ways, the most dominant being either by attempting to model the acoustical properties of a venue algorithmically, or by using an actual digital-sonic 'imprint' of the physical space itself (known as a sound 'impulse' of the space). The latter is achieved by performing a controlled recording of a prescribed sound (e.g. a passage of white or pink noise, or the burst of a balloon) and comparing the recorded result in that venue with the original (the 'dry') signal.

Echo. This is a family of digital effects inspired by the natural phenomenon 'echo'. The only difference with artificial echo is that we can actually control the repetitions (timing, number of repetitions, amplitude), sometimes resulting extremes, often used as novel aesthetic artefacts.

Delay. Delays are again time-based effects. We can have very impressive results with careful usage [and programming] of delay effects (especially when the repetitions are carefully planned to correspond with the musical time (time signatures and tempi). Known popular musicians that have mastered delay technology are musicians David Gilmour (and his famous delayed guitar sounds for Pink Floyd) and Edge of U2 (*when the streets have no name* is a very good example of creative delay programming).

Pitch shifting. Pitch shifters affect the pitch of a recording. This is very handy with some modern music-making technologies that are loop-based. With modern, sophisticated algorithms we can affect the pitch of a recording without affecting the time (if that's what is wanted). You can use pitch-shifters for changing the pitch of an entire recording or use even more sophisticated algorithms for correcting out-of-tune singing (an ubiquitous technology for the recording of musically 'challenged' pop-idols, it seems). Sophisticated, new-technology pitch shifting software are sensitive to formant (see chapter XXXXXX) shifting in order to enable us to perform more realistic correction of sung performances.

Chorus. Chorus effects are based on both pitch-shifting as well as time processing. This is somewhat similar to what occurs within an actual choir. You cannot possibly have two singers sing in unison producing exactly the same sound, in the exact same time. This is what the chorus effect is trying to simulate. This effect has been very successful with guitar and piano sounds.

Harmonization. Harmonizers are very similar to pitch-shifters. The difference is that harmonisers output the original recording mixed with additional voices as well. Such devices (or, again, algorithms) can be programmed in different ways (number of additional voices, how many harmony parts, mix levels), but they can also be programmed to either work within a predefined chord, or as dynamic harmonizers either triggered by performers in

real-time (e.g. using a MIDI keyboard or controller), or with a fluctuating harmonic envelope programmed in advance (for a live performance) or post hoc (during editing vocal performances on a computer).

Compression

Audio compression is a form of data compression designed to reduce the size of audio data files. Audio compression should not be confused with the compression effect (part of dynamics processing) described above. Streams of data (in our case digital audio related data) are 'passed through' audio compression algorithms that have been designed in order to render the original datasets into lighter (i.e. smaller) datasets. The two main categories of compression are: first where the compression process is perfectly reversible (i.e. where we can re-create the exact, unchanged, original dataset from the 'lighter' compressed one); this is called *lossless* compression. Second, compression where one is not in a position to recreate the original dataset (i.e. as it was exactly past A/D conversion) is called *lossy* compression. The most popular audio compression algorithm to date is the MP3 compression algorithm (also known as MPEG layer 3). What is important for the present discussion is that MP3 is a *lossy* compression format. This means that in order to reduce file-size we are getting rid of information that *cannot* be retrieved at a later stage. Known *lossless* compression formats are FLAC and Apple Lossless formats.

<u>educators / practitioners</u>: as with earlier suggestions, what is important to safeguard is the systematic approach to recording, capturing, editing and storing audio recordings. Where there is no real logistical burden to employ lossy compression (e.g. a singing school that performs digital recordings of all taught sessions, in multiple rooms, therefore needing vast amount of digital storage), it would be advisable to archive recordings either uncompressed or, at least, using a lossless compression algorithm. When compression is unavoidable, it would be advisable to perform it at 128 kilobytes per second (Kbps) the least (this is called the 'compression rate' and it essentially determines the quality of the resulting compressed file, which is either strongly or perfectly correlated to its final size). This will ensure that practitioners can listen to reference recordings and assess singing development longitudinally without experiencing problems and without audible artefacts in the digital files.

researchers / scientists: once again within scientific research, in tandem with a systematic approach, we need to ensure that the datasets are in their purest form and that we maintain (and monitor) their integrity. Researchers therefore need to ensure that past the appropriate microphone, and the transparent microphone pre-amplifier sits a high quality, high dynamic range, A/D converter. It is important to note that not all A/D converters are built the same and therefore not all A/D converters perform conversion of the same quality. This is why the previously mentioned argument that 'digital is good quality' is somewhat flawed, as conversion of an analogue signal to a stream of digital information is not a deterministic task. This is why professional recording studios have to perform major investments in their A/D (as well as D/A sections, see below). Past the A/D conversion, researchers should avoid data compression. Finally, the affordances that digital audio introduces and the opportunities for fast, sometimes instantaneous, processing of digital files, harbours the threat of mishandling of those files and the erratic monitoring of their different versions at different junctures. This was not a common threat when people had to utilise physical tape, the manipulation of which could almost certainly not occur on a whim. Within the digital domain, researchers and practitioners need to introduce strict systems on versioning, tagging, file naming, storing and archiving in order to avoid the risk of losing their work, or jeopardizing the integrity of their data.

artists / producers: regardless of the liberty to experiment with available technologies within the digital domain, artists and producers also need to be systematic in order to ensure that their work is safe, but also replicable. Lost work is something that most people have suffered. Another 'ailment of these times' is perhaps also related to the ease with which producers and artists can now achieve novel sonic 'products'. Often, people lose track of the different steps and processes involved in achieving a particular sound or effect and are left unable to replicate the previous steps taken. Evolution and versioning in this context is something that the field could benefit from, perhaps by adopting useful principles from the field of computer science and software engineering. This is seen to have artistic value, but it could also be seen as valuable for the safeguarding of intellectual property.

all three above mentioned groups: as explained earlier, within the digital domain we are dealing with 'information'. In order to perform any action within the digital domain, valid information is essential, but having any kind of information is actually vital. This leads to one final notion of this complex world that is presented in a somewhat naive (or 'accessible') way within this chapter; *clipping*. This occurs when the A/D converter is overloaded with a signal which would need to be converted to a numerical value greater than the converter could handle at a given setting. This, to some, might sound similar to the phenomenon of saturation, overdrive, and/or distortion that one might experience with analog circuitry (which many people actually try to achieve intentionally in order to introduce character or warmth). Within the digital domain, though, this introduces unpleasant audible artefacts, and, unfortunately, complete loss of data past the clipping point. It is somewhat disheartening to witness researchers presenting sound waveform visualisations that show clipping artefacts (easily identified as straight horizontal lines at 0 dBFS) even within published research papers and/or presentation slides at conferences.

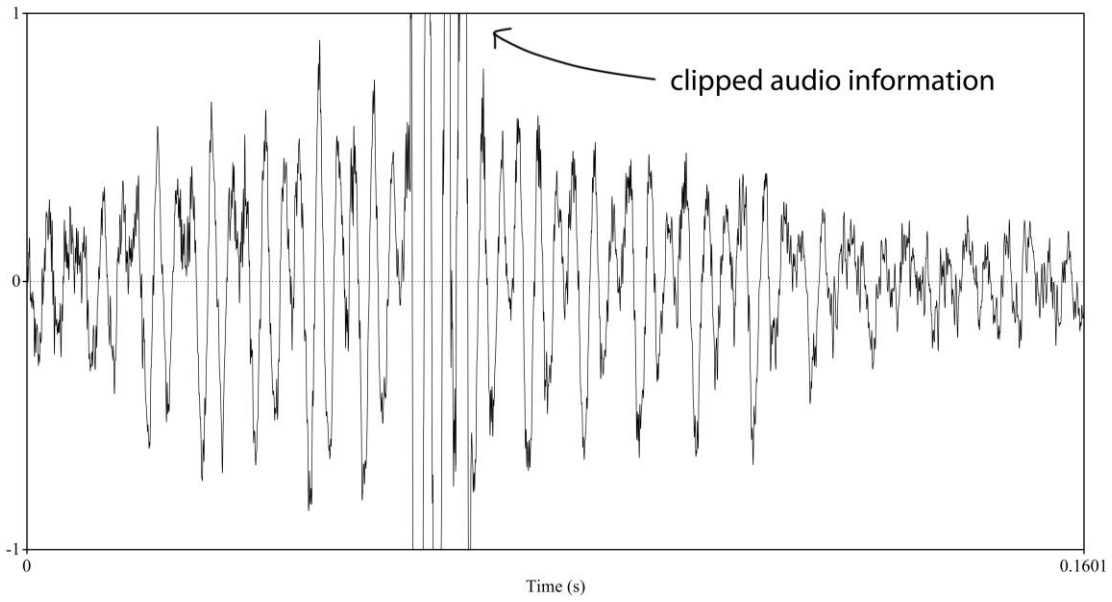


figure 10: a waveform representation of digital audio data where clipping is visible

Outtro

Music Technology is a continually evolving concept. Historically, having access to specific equipment could really make a difference, both in the recording world but also in the production and composing worlds. The things that people could do in the old RCA studios,

the innovatory techniques that the Beatles used for their albums, the music of Vangelis were - to a great extent - connected to the technologies that were available and at hand. When Chariots of Fire was produced, the places in the world that could be used for the production of the specific sounds that were used for the album could be counted on the fingers of one hand. Today, believe it or not, you could produce those sounds with a handful of free software plugins. Gaining the knowledge to be able to create those sounds and be able to put them together in a composition is a different thing, a different art, a different craft, a different science.

In the singing studio, having access to novel technology that allows real-time analyses of singing output and high resolution digital recording of each lesson offers wonderful new opportunities. The effective use of the available technologies, though, is not self-evident. Similarly, in the world of research, novel tools allow us to perform complex analyses in real time; some of these analyses could have taken months to complete even two decades ago. Analytical software can perform numerous analyses of complex datasets regardless of whether certain analyses are meaningful and/or in violation of basic scientific principles. This introduces new challenges for the practitioners, researchers and digital 'natives', challenges that we are not necessarily fully equipped to deal with. This introduces the need for new, systematic educational praxes that are context sensitive and in line with what tools and affordances are now available to us. The future 'studio' beit recording- production- educational- or research- is almost certainly going to be centred on computers, with only the ubiquitous 'microphone' out of 'the box'. Future generations of recordists, producers, musicians, practitioners, researchers, scientists and enthusiasts are likely to benefit from 'sound education' in the systematic acquisition and handling as well as the critical processing and assessment of digital information.

List of References

Berliner, E. (1887). *U.S. Patent No. 372786A*. Washington, DC: U.S. Patent and Trademark Office. Retrieved from https://patents.google.com/patent/US372786A/en

Edison, T. A. (1878). *U.S. Patent No. 200521*. Washington, DC: U.S. Patent and Trademark Office.

Everest, F. A. (2001). *The master handbook of acoustics* (4th ed). New York: McGraw-Hill.

Himonides, E. (2012). The misunderstanding of Music-Technology-Education: A Meta-perspective. In G. McPherson and G. F. Welch (Eds.), *The Oxford Handbook of Music Education* (Vol. 2, pp. 433–456). New York: Oxford University Press.

Howard, D. M., and Angus, J. (2016). *Acoustics and psychoacoustics* (5th edition). New York ; London: Routledge.

Malloch, S., and Trevarthen, C. (Eds.). (2010). *Communicative Musicality: Exploring the basis of human companionship*. Oxford: OUP Oxford.

Mithen, S. (2006). *The singing neanderthals: the origin of music, language, mind and body*. London: Phoenix.

O'Connor, R. (2018, January 3). Vinyl sales increased again in 2017. Retrieved February 11, 2018, from http://www.independent.co.uk/arts-entertainment/music/news/vinyl-sales-2017-bpi-music-report-uk-industry-ed-sheeran-rag-n-bone-man-amy-winehouse-most-popular-a8140261.html

Švec, J. G., and Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, *19*(4), 356–368.

The Physics Classroom. (n.d.). Retrieved February 10, 2018, from http://www.physicsclassroom.com/

Welch, G. F. (2005a). Singing as communication. In D. Miell, R. MacDonald, and D. J. Hargreaves (Eds.), *Musical communication* (pp. 239–259). New York: Oxford University Press.

Welch, G. F. (2005b). We are musical. *International Journal of Music Education*, *23*(2), 117–120.

White, P. (2003). *Basic microphones*. London: SMT.