Tampere University of Applied Sciences

Production of an Audiobook

Recording, Editing and Mastering

Juha Seppänen

BACHELOR'S THESIS
April 2022

Degree Programme in Media and Arts
Music Production

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Degree Programme in Media and Arts
Music Production

SEPPÄNEN, JUHA:
Production of an Audiobook: Recording, Editing and Mastering

Bachelor's thesis 72 pages
April 2022

Audiobooks have grown in popularity during the last decade. The medium for this form of literature has transformed from CDs to streaming, which together with smartphones and widely available internet has made audiobooks more accessible and popular than ever.

The goal of this thesis was to collect best practices for how to record, edit and master a quality audiobook. As many narrators record from home, there are many aspects in the process they should understand – even if they were not the final engineers of the recordings. The thesis goes through the process in chronological order, starting with the preparing the recording space and ending with the delivery of the mastered files.

The bigger purpose of the thesis was to gather simplified but exact and relevant information for a narrator who may not already be aware of the technical aspects of audiobook production.

The theoretical part studies the three parts presented in the thesis title: recording, editing, and mastering. Different methods and pieces of equipment are compared with the subtext of how one can achieve a quality result, even with a smaller budget.

The practical part of this thesis was an empirical research on the three aspects of audiobook production for commercial distribution in the Finnish market from the viewpoint of a narrator-engineer. The recording process of one audiobook is described at the end of the thesis.

Key words: audiobook, narration, recording, editing, mastering

CONTENTS

ABBREVIATIONS AND TERMS

Audio interface	A device that connects for example microphones to the computer, often via USB protocol.
DAW	Digital Audio Workstation. A software used to process audio.
DIY	Do-It-Yourself.
Editing	The process after recording. In editing, for example, unwanted sounds are cut, and pauses are extended or shortened according to specifications.
EQ	Equalizer. Either a plugin or a hardware unit used to control the gain of a range of frequencies.
High-Pass Filter	A common function of an equalizer. A high-pass filter removes frequencies below a set cut-off point.
HVAC	Heating, Ventilation and Air Conditioning. Common sources of constant background noise.
LDC	Large Diaphragm Condenser. The most common type of microphone for studio recording vocals
LUFS	Loudness Units relative to Full Scale. A measurement of loudness over a period of time.
Mastering	The stage of production where a recorded and edited audiobook is prepared for delivery. The master is from which all the formats of audiobook are made from, whether they are downloadable or streamable.
Owens-Corning 703	A common fiberglass insulation board which is often used to build acoustic absorbers.
SSD	Solid-State Disk. A mass-storage device that has no moving parts, being effectively silent.
USB	Universal Serial Bus. A common protocol to transfer data between devices, for example between an audio interface and a computer.

1 INTRODUCTION

This thesis studies audiobook production. It focuses on the technical aspects of the three stages each book goes through: recording, editing, and mastering. Many of the presented practices are applicable for podcast production and other voice recording as well, which may be of interest to some readers.

The thesis starts with considerations for the recording space. Since in an audiobook, there is no music to mask the fact if the narration was recorded in a very reverberant space. The goal of narration recording is to capture a clean, natural, and compelling human voice. A lot of emphasis is given to the recording process, since that is where most of the work happens. The recording space, signal chain and narration must work well for editing and mastering to work. When recording has been done well, editing and mastering are merely the final touches.

The recording process starts with preparing the room. Two of the most important requirements for the space are low noise and little reverberation. Recording equipment, such as microphones, are studied as well. Finally, best practices for a noise and distraction free narration are studied as well.

The next stages are proof listening and editing. The proof listener makes sure there are no errors in the narration, such as faulty words, disturbing noises, or technical issues. In editing, faulty words are corrected, unwanted sounds are removed, breaks are added, and the book is split in chapters and other parts.

Final step is mastering, where redundant frequencies are removed, noise floor is reduced, sibilant sounds are controlled and finally all the files are brought to the same loudness level to prepare for delivery.

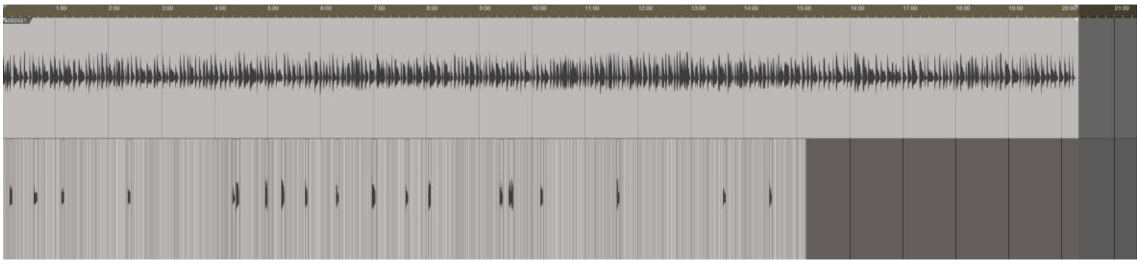
2 RECORDING

2.1 Noise

Human speech is not a continuous string of words with no pauses. There are pauses between sentences to make the text enjoyable, just like in a dialogue between two humans. The narrator must also breathe, as they are a human being.

Breathing pauses serve a few purposes. One is that humans must breathe to stay alive (Nursing Times 2018). Speaking also spends more energy than being sedentary, as it is a function of muscles (Clement n.d.). The other purpose is that narration without any pauses would be demanding for the listener and practically impossible to follow. There would be no time to process the information and the listener would quickly tire. When consuming a traditional book, the reader gets to decide when to take a break.

Just like in music there is silence between the notes, in speech there are also breaks between words and sentences. Picture 1 shows the amount of silence present in narration.



PICTURE 1. 20 minutes and 19 seconds of narration with silence on the upper track and without silence on the lower track (Seppänen 2022)

Picture 1 presents two tracks of audio. The one above is a chapter of a book as it was recorded. The track below is the same chapter run through a macro that removes the silence and closes the resulting gaps.

The original narration is about 20 minutes and 19 seconds long. The narration with the noise and silence removed is about 15 minutes and 10 seconds longs. So about 25 % of the narration is practically silence, but something that must be in the finished product for it to be listenable.

If we consider that 25 % of this book that we have one chapter from in Picture 1 is silence, a book that is 13 hours long will have 3 hours and 15 minutes of silence. Of course, there are breaths and mouth noises and possibly other noises in there, but it becomes obvious that it matters how loud the noise in the recording space is. And of course, if the noise is loud enough it can also be heard on top of the narration. Chapter 4.2 studies noise reduction in the mastering stage.

Noise can be split into three categories by type:

- continuous hum and hiss
- random, transient noises
- human noises from the narrator.

The usual sources for continuous hum and hiss include

- lighting
- preamp noise
- computer fans
- microphone’s self-noise
- HVAC or Heating, Ventilation, and Air Conditioning
- electric appliances such as fridge, freezer, and other household appliances.

The random, transient environmental noises include but are not limited to

- pets
- traffic
- neighbours
- other people near the recording space.

Continuous hum and hiss are quite easy to deal with, since it is usually constant and predictable, which make it easy for a software to reduce. Random noises are much more difficult to remove if they are loud enough. And if they happen in the middle of a word, in most cases the whole sentence must be recorded again.

2.1.1 Managing the continuous noise

The recording space should be as quiet as possible. Archtoolbox (2021), a web source of technical and professional practice reference for architects, states that an acceptable level for theatres, concert halls, and recording studios is 25 to 30 dB-A. The A stands for A-weighting, which imitates human hearing. A-weighting gives less emphasis to low frequencies, to which human hearing is less sensitive to (Neumann n.d.)

Building a recording space that has a noise level below 30 dB-A takes a lot of money and is out of reach for many people. According to Vinnie (2021), building a sound proofed room is building a room within a room. With less monetary resources one can build acoustic elements themselves and achieve an acceptable result.

Because narration in general is not as loud as singing or maybe even character acting, the microphone needs more gain, and noise easily becomes a matter of concern. One of the first sources of noise to control is the category continuous hum and hiss. HVAC or Heating, Ventilation, and Air Conditioning) is often the culprit for that kind of noise.

Probably most buildings in Finland have a heat distribution system based on the circulation of water (Samula n.d.). Each room is equipped with at least one radiator, which is part of the water loop. Basically, the only thing a person can try to control the noise emitted by a radiator is to close the tap that allows the water to flow into the radiator. This may lower the noise, or it might not. Covering a radiator is not advised and even forbidden, since it causes a fire hazard. Luckily, the noise emitted by radiators is usually extremely low and of no great concern.

Ventilation is a much bigger concern. There are supply vents and return vents: supply vents blow air into the room and return vents suck the air from each room and send it back into the system. (Summers & Zim’s, Inc. n.d.) The exact configuration and noise levels vary from space to space, but ventilation is something that needs a bit of thought.

A quick trick that can work is to block the vents. This may upset the distribution somewhat and affect other rooms or apartments that are part of the same system and is probably something the householder would not approve. But done in moderation only for short periods of time, meaning a couple of hours at maximum, blocking the vents probably does not have consequences – apart from a cleaner recording.

Sometimes lighting makes audible noise too. Fluorescent lights sometimes make a buzzing sound, that for some people may be irritating, and if loud enough, it can add to the noise present in the recording. Fluorescent lights have a ballast that regulates the current running through the light (Enoch 2019). Most residential fixtures use a magnetic ballast, which runs at 60 Hz. The solution is to replace the magnetic ballast with an electronic ballast, which runs at 20 000 to 40 000 Hz, which should end all noise. (Enoch 2019.)

Of course, one can just turn off the lights. But if the only source of light is directly from a bright computer screen, that can cause eye strain (Warehouse-Lighting.com n.d.). Laptops usually have a quick way to change the brightness with a key command, but separate screens usually require some menu diving. And if one works in the same space in brighter light, they may need to adjust the screen for that situation again. The process can become repetitive and annoying. A better alternative is to get an adjustable floor lamp with an LED bulb, preferably something that can be dimmed to preference.

If one is recording in a space that is not specifically built as a recording booth, chances are that in the same space or at least in the adjacent space there is an appliance that is often or even constantly on. Household appliances such as fridges, freezers, dishwashers, and coffeemakers are some examples.

When recording, it is best to have no appliance running. At least not in the same space. Coffee can brew on a break and dishes and laundry can wait till after the session. Fridges, particularly older models, tend to create all kinds of noises ranging from low-frequency humming to mid-frequency madness akin to a cow's moo. If there is nothing in the fridge that is immediately destroyed when the power is cut off, it is best to turn off the fridge for the duration of the recording. One just needs to remember to turn it back on.

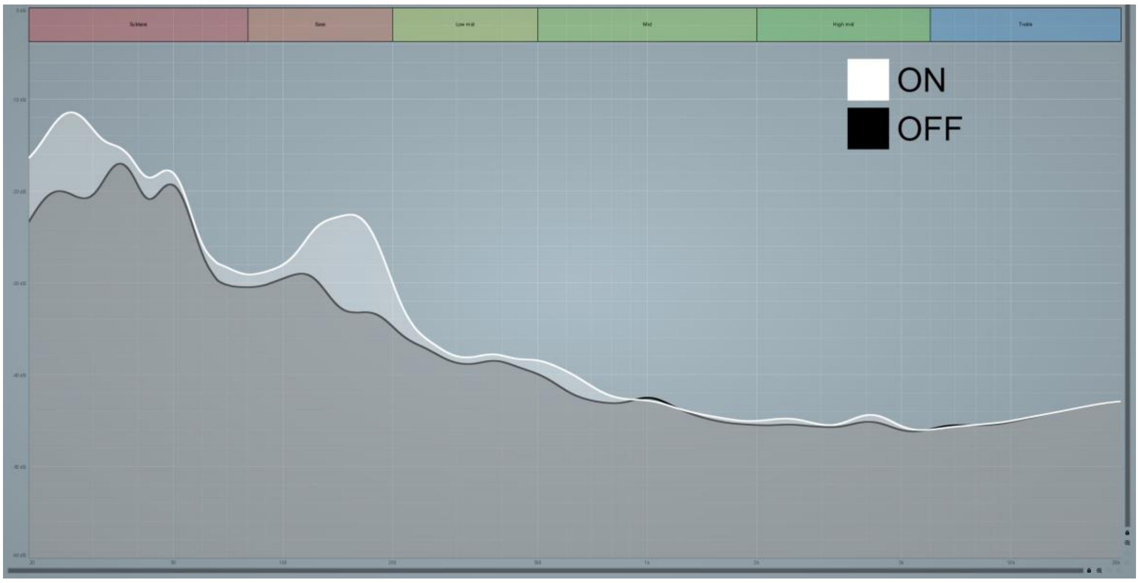
Of less concern than HVAC or household appliances is the noise made by the recording equipment itself. Microphones and preamps make noise, but the levels on any decent gear are so low that environmental noise is much louder. Neumann (n.d.) explains that a microphone's self-noise is usually given in dB-A. The A stands for A-weighting, which imitates human hearing. A-weighting gives less emphasis to low frequencies, to which human hearing is less sensitive to (Neumann n.d.). What is worth noting when shopping for a microphone is that A-weighting results in lower numbers than other measurement standards (Neumann n.d.).

Neumann (n.d.) lists four categories for a microphone's self-noise:

- 11-15 dB-A: Very good. About as low as can be on a small diaphragm condenser microphone.
- 16-19 dB-A: Good enough for most purposes. Can possibly be heard on quiet instruments, but usually unobtrusive.
- 20-23 dB-A: High figure for a studio microphone. Not suitable for anything below speaking level.
- 24 dB-A and above: Unworthy of a studio microphone.

2.1.2 A noise floor test

A noise floor test was done in a studio apartment in a block of flats built in the 1960's. Room noise was recorded with a large diaphragm condenser microphone, the Aston Spirit. Picture 2 shows the noise floor of two different situations. The ON-graph is with all air vents open and a fridge running. The OFF graph is with the air vents blocked and the fridge turned off.



PICTURE 2. Differences in the noise floor in a typical studio apartment with air vents open and closed and a fridge turned either on or off (Seppänen 2022)

The results were averaged over a period of 30 seconds. In the ON-graph, there are two bumps in the lower frequencies: one around 30 Hz and one around 150 Hz. The 30 Hz bump will be cut away by a high-pass filter in mastering, but the 150 Hz bump is right around the fundamental frequencies of both female and male voices. According to Russel (2020), the fundamental frequency for a female voice is generally 200 Hz and 125 Hz for a male voice.

In mastering the level of noise will rise together with the level of narration. Although the noise levels are very similar in these results, the bump around 150 Hz could become a problem. But continuous noise is not such a huge issue, since noise reduction plugins can handle decent noise levels easily. More of a concern are random, transient noises that break concentration and ruin a piece of the recording immediately.

2.1.3 Managing the transient noises

Noise happens. If one lives near other living things, noise is inevitable. If the thump, click, pop or bark happens on a breather, it is easy to replace it with the regular, constant room noise that noise removal plugins manage easily. But if the noise happens during a sentence and is loud enough, the whole sentence needs to be recorded again for the narration to sound consistent.

If one has built a recording space in their own home, chances are that they have other family members, pets, and neighbours. These individuals have their own routines, and they may not care about recording as much the narrator. They cook, they clean, and they watch TV. They also meow to get outside (Becker 2016.).

Modern apartments usually have better soundproofing than older buildings (Knuuttila 2021). But still, rare is the apartment that is built with recording in mind. If one lives in a typical three-room apartment in a block of flats, it is best to choose the room furthest away from the kitchen and the living room to serve as the recording space. Even better if this room does not share a wall with any neighbours.

Family members can usually be reasoned with, so one can ask them to do something quiet when the narrator is recording. They can for example knit a sock while watching TV with headphones. Pets are not so easy to reason with, but the least one can do is to not let them come into the recording space.

Neighbours are a different matter altogether, but communication is the key. Maybe they have a time of day when no one is home, so that time can be used for recording. It does not hurt to ask. Apartment buildings are devilishly hard to silence, though. If only possible, one should avoid living in one if recording regularly.

If there does not seem to be a quiet recording time during the day, the possibility is to record at night when most people are asleep (Özdemir n.d.). That is at least until the narrator rents a separate room or builds their own recording space.

2.2 Acoustics

Acoustics is a convoluted area of physics, and as an exact science is outside of the boundaries of this thesis. As it is a science that requires understanding and delving into the matter, and because the internet is full of advice written by people who themselves may not understand the area, acoustics is often misunderstood and neglected. However, it is difficult to talk about recording without mentioning the acoustics at all.

Reverberation affects the intelligibility of speech (George, Festen & Houtgast 2008). It is also difficult to remove reverb from a recording, so it absolutely must be dealt with before recording (J'vlyn 2017). A space that is intended for recording should be quite dead, which means it has as little reverberation as possible. Deadening a room needs absorbent material, such as rockwool, which is an insulation material very often used in studios.

But often the first type of acoustic treatment people starting up encounter is acoustic foam. Foam is not a very absorbent material, and not suitable for a recording room as the only material. Combined with panels and bass traps made of rockwool they add to the absorption, but on their own they do not do much. Musician and engineer Ethan Winer (2016) presents a table (Table 1) on his website that compares the absorption coefficients of the often-referenced insulation material, Owens-Corning 703, and typical acoustic foam.

TABLE 1. Absorption coefficients of Owens-Corning 703 and a popular brand of sculpted acoustic foam at different frequencies (Winer 2016)

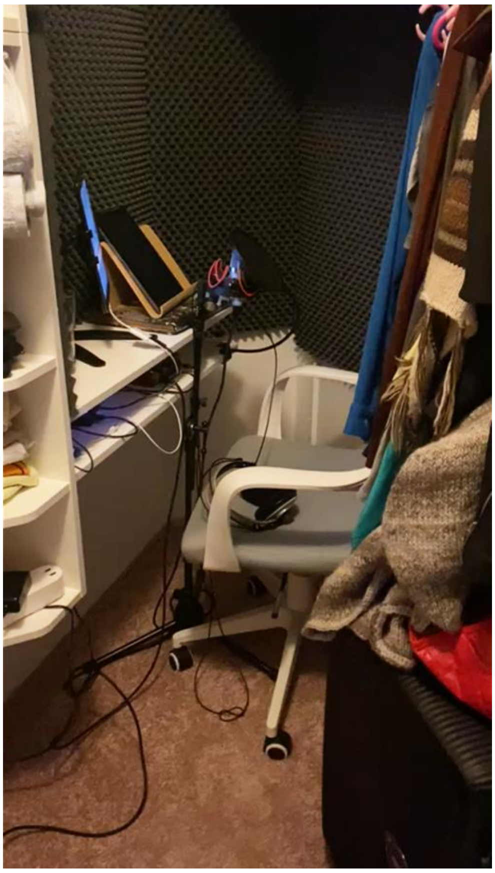
Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Owens-Corning 703	0.17	0.86	1.14	1.07	1.02	0.98
Typical sculpted acoustic foam	0.11	0.30	0.91	1.05	0.99	1.00

The data was gathered from literature published by the manufacturers. All material in the table is five centimetres thick and applied to a wall directly. The table shows that rockwool is more effective at all frequencies, except for the 4000 Hz, where the difference is virtually non-existent for the 703 and acoustic foam. Acoustic foam is most effective, some say only effective, on higher frequencies, and thus should not be used as the only material. The efficiency and resulting sound also depend a bit on the size of the room.

Rockwool panels can be built DIY. Not only do they work better than many commercial solutions, but they also end up being much cheaper. There is the threshold of DIY, but one can always contact a local carpenter with their needs.

Narration does not require a big space. If there is a corner of the apartment the narrator lives in, they can dedicate one corner to their job and construct a small vocal booth. This idea is present in the background when people record in their walk-in closets, but usually they lack any proper acoustic treatment, which results in a build-up of mid and low-mid frequencies combined with reverberation. Often that kind of sound is referred to as boxy.

Narrator Annica Milán has a dual-use walk-in closet in which she records her narration, and she has made it work well. In addition to the clothes she keeps there, she also has acoustic foam on all the walls around the recording spot and also on the shelves (Milán 2022). The core of Milán’s setup can be seen in Picture 3.



PICTURE 3. Annica Milán’s recording setup in February 2022 (Milán 2022)

Together with Swann Studio’s chief audio engineer, Benjamin Blaabjerg, Milán has been able to make the closet work as a vocal booth (Milán 2022). Even if the booth does not have any heavier material such as fiberglass akin to Owens Corning 703, the recordings she makes there have virtually no reverberation or mid-frequency build-up. It is a small space with not much space to move around, but narrators do not really move around like singers sometimes do.

Milán tells that one of the biggest challenges in audiobook recording were actually the acoustic properties of the place she used to record her first books. It was an apartment from the 70’s right in the centre of Tampere, and due to its location, soundproofing, and neighbours, it was not a distraction-free environment. She realised she still had many parts of the series Jääkansan tarina to record and decided to move to a different place. According to Milán, the new place was so quiet that it is difficult to find any complains about her job today. (Milán 2022.)

2.3 Microphone

Together with acoustic treatment, the microphone is where most of the budget should go. Microphone is a piece of equipment that can be heard, and if the sound quality is too far from ideal, it cannot be salvaged in mastering. That does not necessarily mean that the microphone needs to be most expensive one can buy, since in 2022, there are many options for smaller budgets as well.

Microphones come in three main categories depending on the mechanism how they capture sound: dynamic, condenser, and ribbon. The ribbon type microphones are delicate and phantom power may even break some models (Royer Labs, n.d.), so condenser and dynamic microphones are more recommendable in that regard.

2.3.1 Condenser

According to Teach Me Audio’s article (2020), condenser microphones use a pair of charged metal plates to reconstruct a sound. One of the plates is fixed – this one is called the backplate. The other one moves – this one is called the diaphragm. When sound hits the diaphragm, it moves, and the distance between the backplate and the diaphragm changes. This movement produces a change in capacitance, which produces the electrical signal that corresponds to the sound. Figure 1 displays a diagram of a condenser microphone’s mechanism. (Teach Me Audio 2020.)

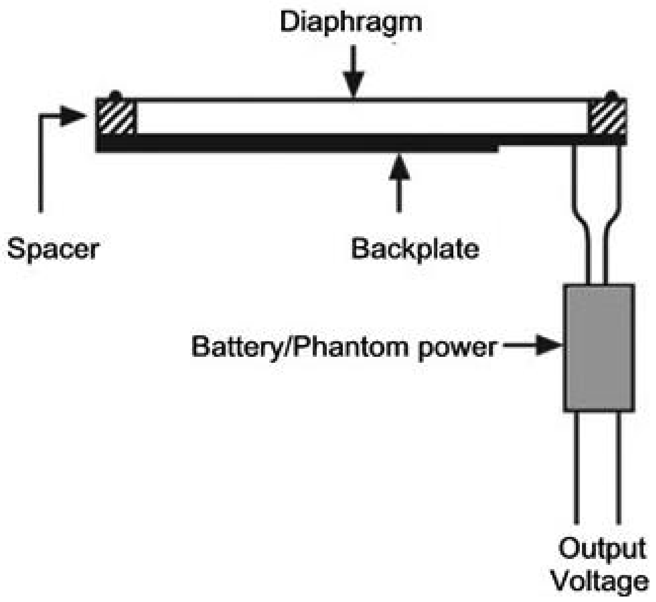


FIGURE 1. How a diaphragm and backplate create a capacitor (Teach Me Audio 2020)

Condenser microphones need an electric current to work. This current is called phantom power, and it is sent from the preamp via the microphone cable to the microphone. +48 V DC is the most common standard for phantom power, but there are others, such as +24 V DC and +18 V DC. Not all microphones require the full +48 V DC, and they have circuitry to manage the voltage they use. (Fox n.d.)

Because of the lighter assembly in condenser microphones compared to that of dynamic ones, condensers are more sensitive at capturing higher frequencies. With a condenser design, it is also easier to achieve a flat and extended frequency response. (Teach Me Audio 2020.)

Since the sonic goal of narration recording is to capture a natural human voice, a condenser microphone is the preferred choice. Not all condensers are equal though, and they can be found in all price brackets, starting from around 100 € up to around 10 000 €. Brauner VM1S is perhaps the most expensive microphone a consumer can buy, but a microphone with that kind of capabilities and price tag to match is not feasible or sensible to use for narration recording.

Sensible choices that are often recommended for narration, voice-overs, and vocals include:

- Rode NT1-A, 179 €
- Audio-Technica AT4040, 399 €
- AKG C414 XLS, 739 €
- Neumann TLM 103, 949 €
- Sennheiser MKH 416 P48, 1019 €.

Prices were taken from music equipment retailer Thomann’s web store on the 2nd of February 2022. They include a 24 % VAT and represent their usual price.

All these recommendations are large diaphragm condensers or LDCs, except for the Sennheiser. The MKH 416 is a shotgun microphone, which means it is a very directional microphone and originally intended for outdoor broadcast use (Sennheiser n.d.). However, its good off-axis rejection means it has a slight advantage at reducing unwanted noises compared to LDCs. The downside is that the sweet spot is smaller compared to LDCs, and if the narrator goes outside the spot, the sound loses its high frequency content fast. The sweet spot and polar patterns are discussed more in chapter 2.9.

2.3.2 Dynamic

Teach Me Audio (2020) explains that dynamic microphones work on an electromagnetic principle. Sound waves hit a metallic diaphragm that is attached to a coil of wire. The diaphragm vibrates the coil in response to the sound, and a magnet that is positioned inside the coil produces a magnetic field. The motion of the coil in the magnetic field generates the electrical signal, which can then be converted back into sound. (Teach Me Audio 2020.) Dynamic microphones do not use phantom power. Figure 2 shows the inner workings of a dynamic microphone.

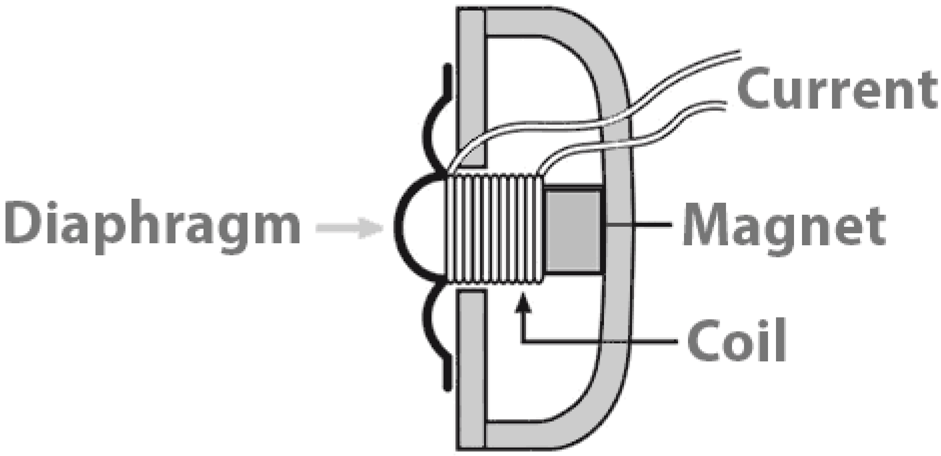


FIGURE 2. The inner workings of a dynamic microphone capsule (Teach Me Audio 2020)

Because the assembly in a dynamic microphone is heavier than in a condenser microphone, dynamic ones do not pick up high frequencies or transients as well (Teach Me Audio 2020). This limits the details a dynamic microphone can reproduce. This is not necessarily always bad, as some narrators might have lots of mouth clicks in their narration. A dynamic microphone can dampen those sounds because dynamic microphones are generally less sensitive to higher frequencies (Teach Me Audio 2020). Then again, mouth clicks can be controlled in post-processing, so it makes sense to record with a condenser and remove the offending sounds later.

The downside of dynamic microphones can also be their upside. While they are not as sensitive and natural sounding as their phantom powered counterparts, they are also less sensitive to the environment they are used in. In a less than ideal sounding studio, which is often any home studio, a dynamic microphone may alleviate the issue of the room (Osburn 2020). They do not pick as much ambient noise or room reverb as condenser microphones. Reverb is difficult to remove from a recording (J'vlyn 2017). So, if one is recording in a less-treated space, it is worth to try a dynamic microphone.

While condenser microphones are the go-to-choice for their natural and detailed reproduction of the human voice, there are dynamic microphones that can do the task simply fine. They may not sound as bright as condensers out of the box but can made to work with some processing. And it always depends a bit on the voice as well.

Not just any dynamic microphone is equally suited for narration, however. What one should avoid are so called stage microphones, such as the ubiquitous Shure SM58. The frequency response chart of SM58 is shown in Figure 3. The manual from Shure (n.d.) says the frequency response is from 50 Hz to 15 000 Hz. The frequency and transient response of those microphones is well suited for live use, but in the studio, it becomes clear that they do not reproduce the human voice nearly as naturally as desired.

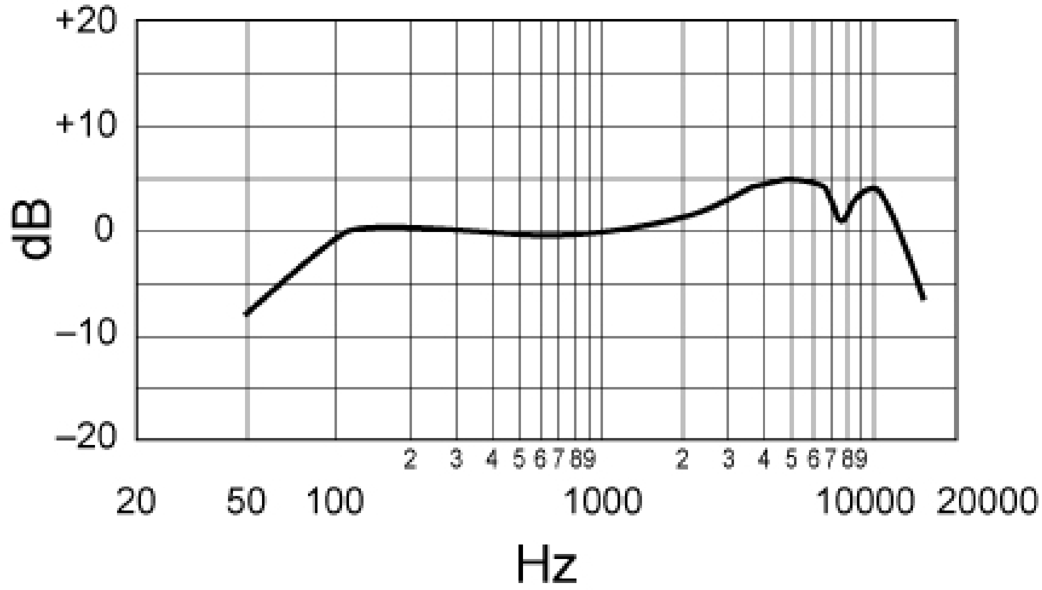


FIGURE 3. Frequency response of Shure SM58 (Shure n.d.)

The dynamic microphones that one should consider are so-called broadcast microphones. Some recommendations are Shure SM 7 B, and Electro-Voice RE20. The first mentioned goes for around 350 €, while the latter goes for around 550 €. The prices were taken from music equipment retailer Thomann’s webstore on the 8th of February 2022. Particularly the SM 7 B has gained traction among podcasters and streamers.

These microphones require lots of gain for speech. Cheaper audio interfaces, like the common Focusrite Scarlett 2i2, cannot supply enough gain for these microphones to be usable. According to Focusrite's Scarlett 2i2 manual (n.d.), the maximum gain the 2i2's preamps provide is 46 dBs. According to Shure (2021), the SM 7 B needs at least 60 dBs of gain. The Scarlett 2i2, like many audio interfaces, are made for condenser microphones that have a much hotter output level.

The solution is a preamp with lots of gain, for example the Grace Design M101. According to Grace Design (n.d.), the maximum gain the M101 can supply is 75 dB, which is enough for the Shure SM 7 B. However, the M101's MSRP is 925 $ (Sweetwater 2022), and thus may be out of reach for someone who is starting out.

What is also worth considering is an inline preamp, sometimes called mic-activator. These preamps are little units that take the unused phantom power from a preamp and convert that into additional gain for the dynamic microphone. There are many models on the market in 2022, and one of them is the TritonAudio FetHead.

The FetHead is a such a small inline preamp that it becomes a part of the microphone cable, dropping the need to buy another cable. The FetHead adds 27 dB of gain, and costs about 69 € directly from the manufacturer (TritonAudio n.d.).

2.3.3 USB microphones

In the past decade, the market has seen many new USB microphones. They are condenser microphones that do not require an audio interface to connect to a computer. Instead, the electronics are built into the microphone, dropping the need for an interface.

USB-microphones can supply adequate sound quality for some purposes. One example could be a conference call where the sound quality is compressed to save bandwidth (Katz 2020). But when it comes to microphones, with a bigger price tag usually comes better sound quality. As USB microphones need to have the preamp and converters built in while keeping an attractive price tag, corners must be cut somewhere (Wreglesworth n.d.). Also, because the components are built in, the end user is quite stuck with them. With separate signal chain components, it is easier to change just one of them.

2.3.4 Pop filter

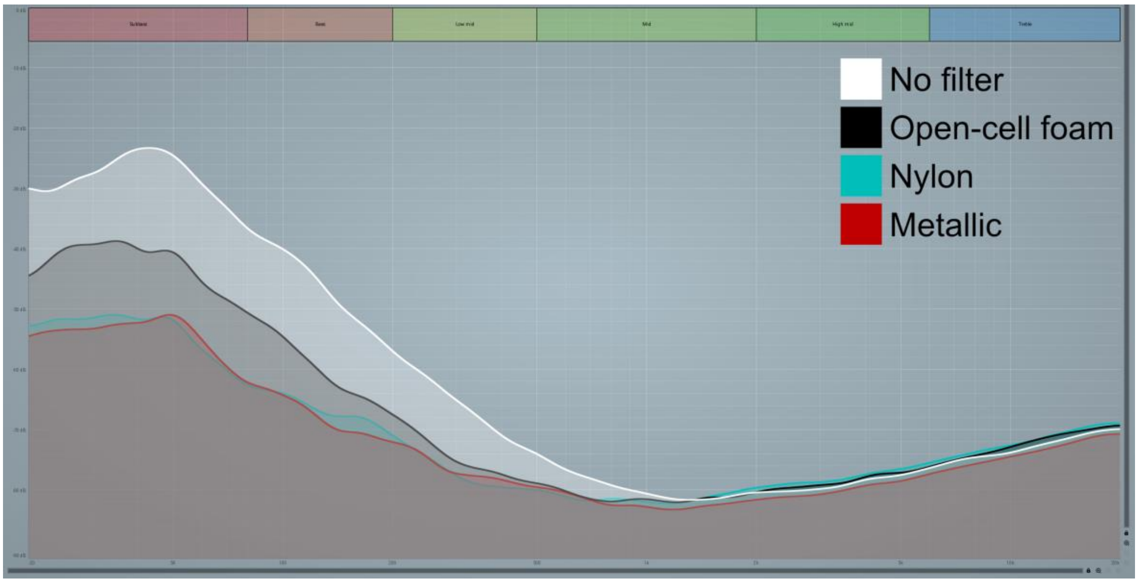
Pop filter is an essential studio tool. Its purpose is to dampen aspirated plosive sounds, such as the first “p” in the English word “pill”, so the burst of air in those sounds does not reach the microphone capsule (Harman 2021). The burst of air appears generally as an under 200 Hz bump.

Probably the most common type of pop filter is a piece of woven nylon stretched over a circular plastic frame. The frame is attached to an adjustable goose neck, which can be attached to a microphone stand. There are also pop filters in which the screen is made of metal or open-cell foam.

The pop filter must be acoustically transparent. Particularly microphone windscreens are not recommended for studio use since they can dampen high frequencies too much (Audio Technica n.d.) They are thicker and made for outdoor use, where wind is often an issue – hence the name. The advantage windscreens have, though, is that as they are tightly wrapped around the microphone, they obstruct the narrator’s view as little as possible. So, it is a matter of weighing the benefits of being able to see the text better and possibly losing some high frequency content. It is hard or impossible to say for certain that a windscreen will negatively affect high frequencies, so again, it is something that needs to be evaluated on a voice-by-voice basis.

For demo purposes, even a sock can function as an emergency pop filter. When narrator Annica Milán had to record a demo for an audiobook company, she was on a cruise ship and thus had no access to a proper studio setup. She explained the situation to the company but borrowed a microphone from her friend and put a sock on the microphone. (Milán 2022.)

Picture 4 shows a test of different pop filters compared to no filter at all using short bursts of air from a can of compressed air. The test was done with a Sennheiser MKH 416 microphone pointed directly at the nozzle of the can from 25 centimetres. A pop filter was placed between the nozzle and the microphone, and the result was recorded and averaged.



PICTURE 4. How the pop filter material affects a burst of air from a can of compressed air. Analyser set to a tilt of 3 dB per octave (Seppänen 2022)

Each white horizontal line presents a gain change of 2 dB. The highest peak without a filter is around 42 Hz. At 42 Hz, the metallic pop filter dampens the signal for about 30 dB. A very similar result is achieved with the nylon filter. The open-cell foam dampens the signal about 16 dB. From about 1 kHz and higher the differences are negligible.

The open-cell foam filter in this test was the Rycote InVision INV-7. The nylon filter was an unbranded gooseneck-type filter. The metallic filter was the Aston SwiftShield. These results seem to show that a metallic pop filter would be the most effective at dampening plosives, but further tests with different brands of pop filters would be needed to draw any definite conclusions.

2.3.5 Microphones afterword

Microphone choice depends a lot on the sound of the narrator and space in which they are recorded (Ciccarelli 2020). One microphone might have too pronounced high frequencies on one voice, while the same microphone could bring a pleasant clarity and brightness to another voice. The ideal situation is if one can test multiple microphones in a single session in the intended recording location. It is also worth to keep in mind that when it comes to equipment, microphone has the biggest impact on how someone sounds.

Even though price can be used as a guide for how good a microphone is, there is a chance that a cheaper microphone works for one voice better than something that costs 3000 euros. It is all about testing as many as possible. Of course, with a 3000 € microphone one probably sounds very good, but something cheaper can be a better match.

2.4 Audio interface and preamp

To connect an XLR microphone to the computer one needs an audio interface. Audio interfaces can have a multitude of features, but from the perspective of recording, two are the most important:

- supplying gain for the microphone to bring the signal to line level
- converting the microphone’s analogue signal to digital for the computer.

Condenser microphones have active circuitry that requires what is called phantom power to work (Henshall 2014). An audio interface generally has any number of preamps from 1 to 8, and these preamps supply the 48 volts of power to the condenser microphone. Almost any modern audio interface has at least one preamp that supplies phantom power.

For preamps, when recording narration or a voice-over, priorities should be transparency and clean gain. Some people may argue that only classic units such as Neve may be used and nothing else comes close, and that everything else is just not worth it (Harris-MacDuff 2021). But in narration and voice-over work, it is best if one uses a clean sounding microphone with a preamp that colours the sound as little as possible – particularly if the narrator is not the person mastering the final product, as they might not be able to stay objective in regard to their own voice. It is much easier to colour a sound in post processing, if at all necessary, than removing a certain quality of sound that has been printed in the file, such as reverb (J'vlyn 2017.).

By clean gain, it is meant that increasing the gain of the preamp does not introduce too much noise into the recording. But preamp noise is something that should be the least of an aspiring narrator's worries. Microphones have more self-noise. It is worth noting that dynamic microphones often do not have a specified noise level, because their noise performance is dependent on the preamp used (Neumann n.d.). And when it comes to order of importance regarding noise, environmental noise places far higher in the list than signal chain noise.

2.5 Computer

Recording happens on a computer. This can be a Mac or a PC. Operating system does not really matter, since there are capable DAWs and simple recording software available on all platforms, Linux included.

The computer can be a desktop or a laptop. Processor performance, or amount of RAM do not matter in recording too much. Other variables, such as the microphone and the audio interface, and stability of drivers may matter mor (Audacity Wiki 2019.).

The computer should be as quiet as possible. Since if one is recording by themselves, the computer is usually quite close to the microphone. What makes the most sound in a computer are the cooling fans and hard drives. (Crucial n.d.)

Desktop computers are rarely built for entirely passive cooling, since the wattage and resulting heat in desktop components is higher than in mobile components (Stone n.d.). Cooling desktop components needs almost constant airflow, or even liquid cooling. Usually at least the processor fan is running. Often there are also the power supply fan, graphics card fan and case fans. When building or buying a computer, a good rule of thumb is that fans larger in diameter produce less noise since they can move the same amount of air with a slower rotation speed (Burke 2011).

Laptops are built different. As said above, the wattage and resulting heat in mobile components is lower. Therefore, on a lower workload a laptop can be effectively silent. And since recording a single track is not demanding, a laptop can stay quiet for a long time. Of course, there are many different manufacturers, and they build and design things differently, but generally laptops can be considered quieter – unless under a big workload. Then the sound can be akin to a jet engine.

Hard drives are another source of noise. Thankfully, the legacy hard drives with spinning disks and a moving arm are on their way out, and solid-state drives (SSD) are increasingly common – especially in laptops. In fact, on the 18th of February 2022, the search term “SSD” returns 610 results in the laptops section of the electronics retailer Verkkokauppa.com’s webstore. The term “HDD” returns 4 results. Only of the four laptops had an HDD, the others had an expansion slot for one. With SSDs, one does not have to worry about their noise at all, since they do not have any moving parts.

The takeaway is that when it comes to single-channel recording, the computer does not need to be excessively powerful. The noise level and ergonomics are more important. Many people already have a computer at home, and it is likely that it will do simply fine – if it is at least reasonably modern.

2.6 Software

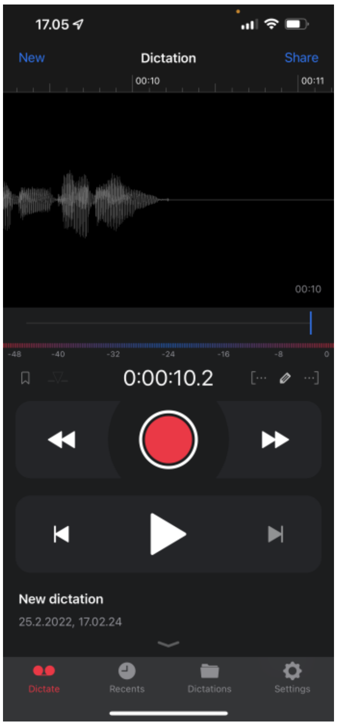
There are three must-have requirements for the recording software. It must be able to record with at least a 44.1 kHz sampling rate and at 16-bit depth, it must have the option for punch-in recording, and it must have crossfades. All of these are standard features, however, so no matter the software of choice it probably has these.

In 2022, there are multiple free pieces of software that can be used for recording. A few freeware are Audacity, Cakewalk by BandLab, and Apple GarageBand. While the latter is technically free, it is available only on Macs, which are not free unless one gets one as a gift. Also, GarageBand does not have automatic fades, but instead one needs to make fades with volume automation. If the narrator must pause often and/or retake something, the process can become very tedious. Apple's Logic Pro, which is sort of an upgrade of GarageBand, can do fades easily.

Audacity is a popular freeware that can do recording. Some say it does not count as a DAW or Digital Audio Workstation, compared to some industry favoured software, but for recording a single mono channel it will do – especially if the narrator does not have to do the final editing and mastering. Narrator Annica Milán (2022) uses a recording software that is very simplified feature-wise.

Some paid for software that with their traditional linear workflow suiting recording are Steinberg Cubase, PreSonus Studio One, Avid ProTools and Cockos Reaper. However, for simple recording purposes these might be considered overkill. They do however offer workflow-related features that make recording easier. For example, in Studio One, with an addon one can make a macro that places a marker at a desired length from the end of a phrase. The length can be for example 3 seconds, which is a standard time of silence at the end of one chapter.

Narrator Annica Milán does not use a computer for recording. She uses two iPads instead, as seen in Picture 3 on page 16. Milán records in a simplified iOS app called Dictate+. The iPhone version of the app's interface is shown in Picture 5 on the next page. Compared to a DAW like Studio One, for example, using Dictate+ is more straightforward.



PICTURE 5. Dictate+ on an iPhone. The iPadOS-version Milán uses looks different, but is almost the same in functionality

2.7 Technology afterword

With all the technology, it is easy to forget what matters the most: The Narrator. The ability to deliver an interesting and pleasing performance matter far more than minuscule differences in a preamp's noise floor or a microphone's ability to capture detail at 13 367 Hz. A good quality microphone paired with a decent audio interface does not break the bank. Especially with the post processing power available today, a simple signal chain will go a long way.

2.8 Basics of human voice

Masterson (2010) explains that humans use the same apparatus as the chimps to speak: lungs, throat, voice box, tongue, and lips. When an individual speaks or sings, they release controlled puffs of air from the lungs through the larynx. When the air travels through the larynx, the vocal cords vibrate, producing sound. The pitch changes according to the tightness or looseness of the vocal cords.

The sound is further shaped by the throat, mouth, tongue, and lips. Professor of cognitive and linguistic science, Philip Lieberman, explains that speech is the most complex motor activity, except for maybe playing the violin or acrobatics. (Masterson 2010.)

Mechanically this motor activity is the basis of narration, and it is what will be transformed by the microphone to electricity, and from electricity to data by the analogue to digital converter. Of course, there is much more to speech and there are areas of science dedicated to studying the different areas. But this thesis focuses on how to turn this mechanical activity to an enjoyable listening experience.

2.9 Audiobook is a book

Audiobook is a book in audio format. It has been narrated, recorded, edited, and mastered by someone. One common narration style is unvoiced solo narration. Unvoiced means that different characters are not given their own voices, and the whole text is read in a natural, straightforward way (Clark 2018).

Some narrators, depending on the book, blend unvoiced and voiced narration, and they give each prominent character a slightly different voice. This may make it easier for the listener to follow the story. If the book has many characters, it might be confusing if all spoke in the same way. Then again, if the book has a lot of different voices, it might be confusing again. So, it all depends on the genre and number of characters. According to Clark (2018), partially voiced solo narration is best suited for general fiction, action, mystery, and fantasy, while unvoiced solo narration is better for romance, non-fiction, thriller, and suspense.

2.10 The process of recording oneself

A possible setup is to have the text on a tablet and record on a laptop. Tablets are handy because they can be easily moved and thanks to the touchscreen, changing the page makes no sound, unless one taps very furiously. It is also worth noting that there are computer mice with practically silent scroll wheels, such as the Logitech MX Master 3 (Harding 2019). The recording happens in software on a laptop, to which the microphone is connected via an audio interface.

The narration of a book starts with an intro in which the author, title, translator, and the name of the narrator are announced. Then the book is read. It is somewhat of a standard to have a three second pause of silence/room tone between different sections of the book.

When the narrator makes a mistake, takes a break, or pauses for any reason, punch-in recording becomes necessary. This means that the narrator can listen to what they said before the break or mistake and then activate recording. The way how punch-in recording is activated differs by the software, but the general idea is the same.

Speaking is a muscle function (Clement n.d.), and the voice can tire after some time. Breaks are necessary and drinking small amounts of water are recommended. Many narrators record a maximum of a few hours at a time and continue later the same day or the next day, depending on how fast the voice recovers. If one records in their own home, they can decide when to record. But if they have paid-for studio time, then they must adhere to it.

Milán’s (2022) typical workday is fragmented. Some days she might record for an hour at 10 o’clock, take a one-hour break and record for two to three hours again. Some days she might record for a few hours at a time and continue the next day. She says that some days her voice just works and sounds better, and she uses that momentum to record for a longer time. This has backfired sometimes, so that the next day her voice has been tired, and she could not record. These days she splits the recording into more even sessions, so she does not risk straining her voice. (Milán 2022.)

Narration may appear simple but requires concentration. One needs to stay on top of the story, support a good energy all the way to the end of a sentence, carry a compelling intonation throughout, and to breathe at proper times. Milán feels that singing and speaking are two very different things and require different techniques. She says that she has a much better singing technique than speaking technique, and that a two-hour live show does not strain her voice as much as two hours of talking can. Milán also draws attention to an issue that many people speak in a lower range than what would be ideal for their voice, Milán herself included. If during a reading session she notices her pitch has started to drop, she raises it higher and focuses on her support. (Milán 2022.)

She also points out that especially women's voice pitch has lowered during the latest centuries. Researcher Cecilia Pemberton from the University of South Australia has written a paper on the matter and the team found that the fundamental frequency of women had dropped by 23 Hz from 1945 to the early 1990's (Robson 2018).

2.11 The narrator as a source of noise

Condenser microphones are very sensitive. They are the preferred choice for recording narration and vocals since they reproduce details well (Ramm n.d.). However, their sensitivity becomes a problem in an untreated space. While the room and the environment are important, also important is how the narrator behaves during recording.

The narrator must remain relatively still when recording. Clothes chafing against the chair can easily be picked by the microphone. Or loose jewellery can rattle, and those kinds of high frequencies generally stick out of the narration easily. Lower frequencies are more easily masked by the narration.

Clothing plays a part in recording. Leather clothing, such as shoes can creak, which is an identifiable, unwanted sound. If possible, shoes should be taken off for recording. A good alternative are comfortable woollen socks. Narrator Annica Milán likes to wear woollen socks when she is recording (Milán 2022). Ideally, when recording, one should wear something made of natural fibers, which are quieter than conventional synthetic fibers such as polyester and acrylic (Doroudiani 2015). Cotton as a material does not make that much high frequency noise when chafing against something. Something like a shell suit would probably be one of the worst things one could wear when recording audiobooks.

The choice of chair matters too. An ideal chair is a chair that does not creak or have many moving parts, especially if they are loose or badly lubricated. A simple dining chair may be better choice than a work chair with complicated mechanisms, especially if they are badly kept. Ergonomics must be thought of too since the narrator needs to sit down for long periods at a time. Standing is a possibility too, but it can become tiresome.

Stomach rumbling is picked up easily too. Sometimes it can be so loud that it can even be heard inside a sentence, and those cannot be removed in editing. Having a small snack during a break can help avoid the issue, since often the source of rumbles is an empty stomach (Bashforth 2021). But narration, like singing, is more difficult with too full a stomach since it impedes the movement of the diaphragm and thus affects airflow. Having too full a stomach also affects energy levels since the body focuses on dissolving the food. This effect can be tested by having a big lunch during the day.

While there are multiple tools that can be used in post processing to mitigate mouth noise, it is good to understand the phenomenon so that it can be controlled already at the source. Mouth noise is sometimes called dry mouth, which can be somewhat confusing because the problem is not lack of saliva. Educator Scott Winstead (My eLearning World 2021) explains that mouth clicks are caused by the saliva becoming stickier and less liquid. Hydration is important and taking small sips of room temperature water during the session can help. It also helps keep the vocal cords hydrated. Important is also to avoid drinking too much water during the session, because this can also lead to mouth clicks.

If the narrator tends to have lots of mouth clicks and water does not seem help, it might be worth considering placing the microphone slightly off-axis, preferably below the mouth (Rempel n.d.). Placing the microphone above the mouth can lead to subconscious craning of the neck, which makes the airway smaller, and the voice may sound strained.

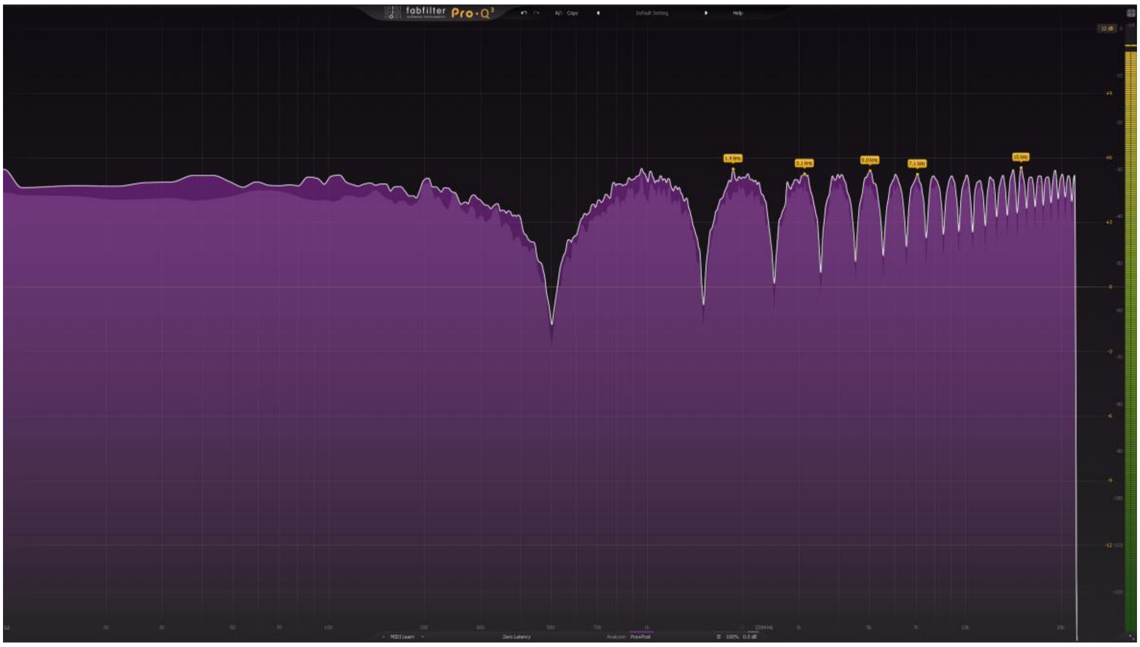
The computer keyboard is also a sound source. Recording software by default starts recording at once when a button is pressed, so a loud keypress can appear at the beginning of every recorded clip. Rubber dome switches, common in laptops, generally trump loud mechanical keyboards in this regard (Computer Lounge 2019.). Fortunately, the keypress appears at the beginning of a clip, so it can be easily edited by adjusting the crossfade for two different clips. Other than pressing a key to start recording, one should not use the keyboard during recording anyway. Of course, this problem does not exist when recording in a studio with a separate control room. But if one is recording by themselves, this is a small thing that should be given some consideration.

A way to avoid the problem of audible keyboard presses completely is to use a pre count feature found in many recording software. Pre count means that the software adds a countdown before it starts recording, and the length of the pre count can be adjusted. This also gives an opportunity for the narrator to breathe.

Narrator Annica Milán (2022) records on an iPad, so she does not need to worry about a loud keyboard. Tapping on a tablet computer that is situated behind a cardioid microphone makes a barely audible, very gentle sound – unless tapped with rage and long fingernails.

Depending on the microphone and how it is mounted, any touching of the microphone stand transmits into the recording quite loud and appears as bass-heavy transients. Also, if the microphone stand is a boom arm fixed to a table or a generic microphone stand that physically touches a table, any accidental (or non-accidental) hitting of the table probably carries on to the recording. A microphone cradle with suspension helps alleviate the risk.

Another thing to consider if the microphone is mounted to a table stand is that the microphone may pick up reflections from the table, screens, or any hard surface. This can lead to a phenomenon called comb filtering, in which the same sound arrives at the microphone, or the listener's ears, at different times with very small delays (Fuston 2021). The delays can be as short as under one millisecond, or up to 20 milliseconds (Fuston 2021). Depending on the delays, some frequencies are reinforced, and others are cancelled out, which leads to the frequency chart looking like a comb, as presented in Picture 6.



PICTURE 6. Comb filtering as a result of two channels of pink noise when one of them is delayed by 1 millisecond. The analyser has been set to 3 dB per octave tilt (Seppänen 2022)

Comb filtering is something to look out for. When recording solo narration, oneself hard surfaces near the microphone are the most common culprit (DPA Microphones 2019).

Self-sourced noises are easy to control since the narrator oversees them. Much harder to control are sounds outside the recording space, such as pets and neighbours. Not everyone has the means to build a floating, soundproof studio in their home. Fortunately, not everyone has to do so. Generally, apartments are fairly good what it comes to soundproofing, maybe excluding older blocks of flats. In those apartments, neighbours easily sound like elephants walking on high heels blowing trumpets all day long.

2.12 Distance from the microphone

Distance from the microphone greatly affects how something sounds (Ciccarelli 2018). If one is too far, the voice appears distant and thin. Terms like thin and weak are often used, and they mainly refer to a limited range of frequencies. The further a sound source is, the less low frequencies it has (Merlot n.d.). Narration recorded too far from the microphone may be described as lacking intimacy and closeness, which refers to the loss of low frequencies.

But too close is also not ideal. As the microphone manufacturer Neumann (n.d.) states, most microphones have something what is called the proximity effect. Proximity effect means an increase in low frequencies the closer the source is to the microphone. This can make the voice sound muddy and negatively affect speech intelligibility. In some voice-over cases, though, the proximity effect can be used to an advantage. Especially when recording commercials, some voice-over artists can get really close to the microphone to sound “bigger” (DelGaudio 2017). The proximity effect is more of a concern for the typical male voice that has a lot of information under 200 Hz. (Neumann n.d.)

Abusing the proximity effect is not recommended in audiobook narration. In books, it is imperative that the listener can understand the text. In Western languages, speech intelligibility is highly dependent on the range of 1 kHz to 4 kHz (DPA Microphones 2021). In Figure 5, DPA Microphones refers to research by French and Steinberg from 1947.

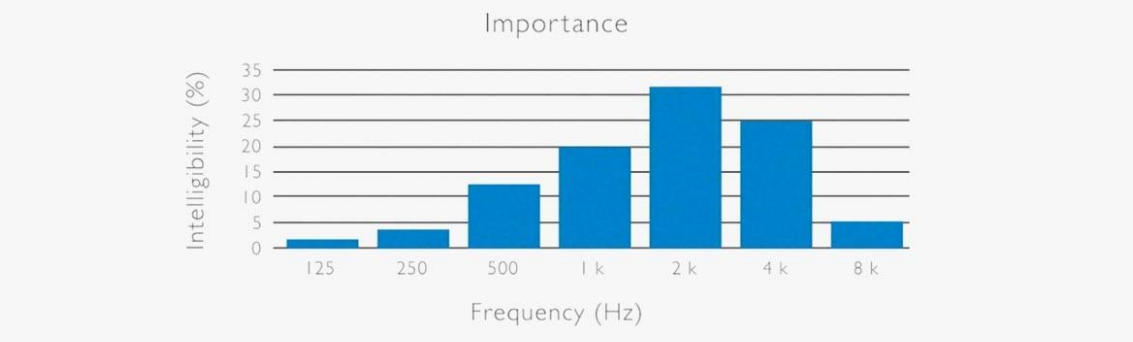


FIGURE 5. Importance of frequencies in Western languages regarding speech intelligibility (French & Steinberg 1947)

The importance of different frequencies can be tested with a high-pass filter. Setting a high pass filter at 500 Hz gives very different results than the same filter at 1 kHz.

There is no single universal distance that works for every microphone. And every microphone’s proximity effect is also different. They are things one needs to test. It is also better to do the test with another person since humans may have difficulties staying objective towards their own voice. But generally, a range between 15 to 30 centimetres is suitable (Merlot n.d.). Podcast editor and sound designer Boudreau (2020) recommends finding the right distance by making the shaka sign, sometimes known as the “call me” sign. The sign is made by extending the thumb and the smallest finger and holding the three middle fingers curled. Then the narrator places the thumb to their chin and the smallest finger to the microphone. That is roughly the proper distance.

2.13 Polar pattern and the sweet spot

For vocal recording, the cardioid polar pattern is a standard (Lewitt 2016). The pattern is shown in Figure 6. They can also have switches for different patterns, but cardioid is generally the most useful – the same pattern is used in a lot of other recording situations too.

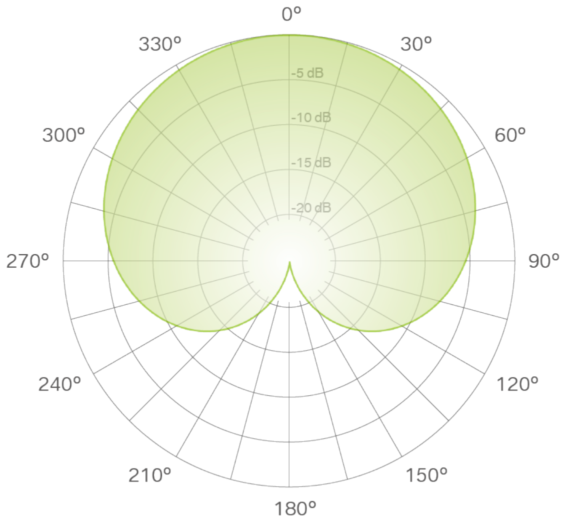


FIGURE 6. Cardioid polar pattern (Lewitt 2016)

In Figure 6, The 12 o'clock position of the circle is the front of the microphone, and the 6 o'clock position is the back. The green area shows from where the microphone capsule picks up sound. It is most sensitive at the front, with least sensitivity in the back. One circle means a 5 dB drop in sensitivity. (Lewitt 2016.)

Because of rejection at the back and somewhat in the sides, cardioid microphones are usable in more echoey rooms than omni-pattern microphones, which pick up sound equally from all sides (Levine 2014). While omni-pattern microphones have the flattest frequency response, they cannot be recommended unless the recording space is very well treated and has no prominent sources of noise.

There are three other polar patterns or types that are worth a mention: supercardioid, hypercardioid and shotgun. The first two are variations of the regular cardioid pattern and have narrower picking angles. Hypercardioid's angle is narrower than of a supercardioid's. Typically, microphones that are specified as super- or hypercardioid differ from regular cardioid not only with their narrower picking angle in the front, but also that they pick up more sound from the back as well. (Peter 2019.) But even if a microphone is specified to have a certain pattern, they are not always as exact as for example in Figure 6. This is also something one needs to test.

Shotgun microphones are the most directional microphones. In these kinds of microphones, the capsule is at the bottom of a few tens of centimetres long tube, which is called an interference tube. Koschak (2016) explains that the tube has several openings along its length, which allow the sound to enter the tube. Sounds that enter the tube from the sides arrive at the capsule at different times. They are out of phase and phase-cancel each other. (Koschak 2016.)

On-axis sounds that enter the tube from the front arrive at the capsule at the same time. They are in-phase, and thus do not cancel out. The result of this design is a narrow, very directional lobe of sound pickup at the front of the tube. Shotgun microphones are characterised by such lobes and are thus sometimes called lobar. (Koschak 2016.)

Even though shotgun microphones are intended for location recording on movie sets and sporting events, for example, with proper technique they can also be used for studio purposes. The proper technique in this case means that the narrator must be vary not to move around much, as especially high frequencies disappear very quickly if not on-axis (DelGaudio 2017). The regular cardioid pattern is more forgiving with movement, but also pickup more sounds from the environment.

2.14 Microphone’s vertical position compared to the mouth

To make narrating easier and thus get a better performance, the microphone must not be placed too high or too low so as the narrator would have to raise or lower their chin to an unnatural angle. This can lead to the voice sounding strained and can make the narrator tire faster.

When starting to find the right vertical alignment for the microphone, the narrator should have their head at a natural, relaxed angle, so that their throat is not constricted at all. It helps if the text they are narrating is found below their eye level. Many narrators have the text on a tablet, which can be easily placed anywhere it is comfortable. When the narrator has found an angle that feels good, the microphone should be placed right at the mouth height.

Mouth level is often fine. But plosive sounds, such as P, are easily exaggerated on this height (Monk 2012). The severity of the problem depends on the properties and placement of the pop filter, which is an absolute necessity when recording vocals up close. Having the microphone right at the mouth level may also lead to increased amounts of mouth clicks.

If the plosive sounds are overpowering, then the microphone should be placed slightly above or below the mouth height. That way most of the air passes above or below the microphone capsule. Below the mouth is preferred because it helps avoid craning the neck.

Also, according to mixing engineer Mike Senior (2009), “higher frequencies tend to beam slightly downwards from the nose and mouth.” So, if a brighter tone is desired, one could try placing the microphone below the mouth.

3 EDITING

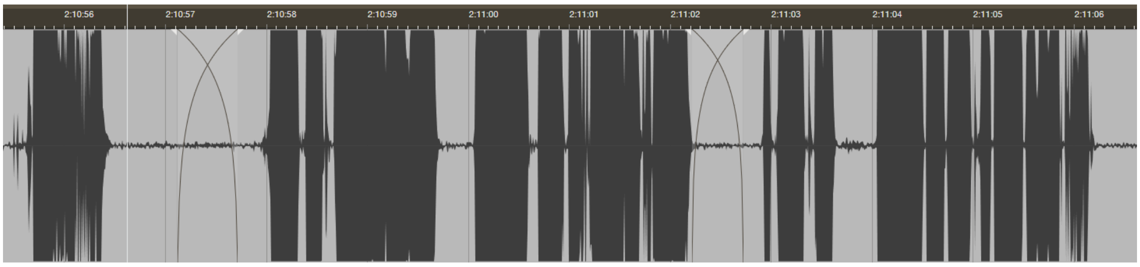
3.1 Proof listening

Sometimes the client requests proof listening (Listening Books 2019). In this process the proof listener listens to the whole book and writes down if there are any errors such as wrong words, loud noises, or technical issues. The proof listener writes down the timecode for the error, the page number, file number, the context around the error and what the error itself is. All these help the narrator find the right spot fast, as the files are usually tens of minutes in length.

Searching for a single word could take some time, if the narrator did not make any markers in their recording software's marker track – if it has one. Most mature DAWs, such as Cubase, Logic Pro, Studio One, Reaper, and ProTools have a marker track or even multiple.

The narrator records the whole sentence again. It is important to record the whole bit since it is practically impossible to say just one word with the same intonation so that it sounds natural with the rest of the original sentence.

Then the editor or proof listener cuts away the original take and replaces it with the new one. Attention needs to be paid to the crossfades around the clip and the volume of the new clip, since the new sentence is probably of different length and slightly different volume. If the new sentence is significantly shorter or longer than the new one, the rest of the clips must be adjusted to shorten or lengthen the new gap. Picture 7 shows a usual case where one sentence has been re-recorded and replaced with a new one.

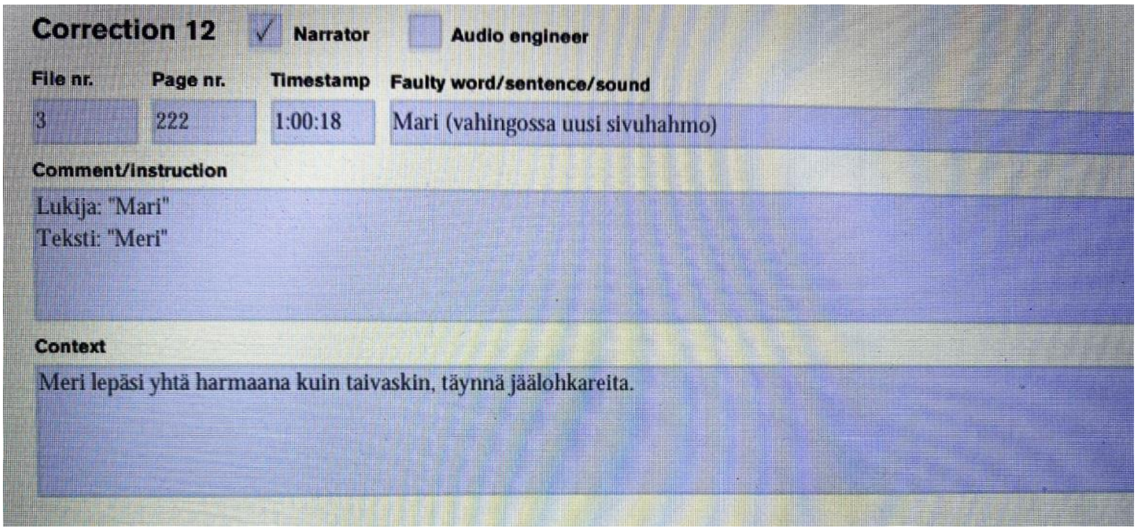


PICTURE 7. A sentence that has been replaced by a fixed take from the narrator. Data zoomed in Studio One to show the noise floor (Seppänen 2022)

Even if the noise is in the mastering stage reduced to an extremely low, hardly perceivable level, it is important to know that when crossfading two sounds the noise should be constant in volume. A common problem with linear crossfades is that the volume may drop a bit in the middle of the fade, which makes it sound not continuous (Audacity n.d.).

The crossfades in Picture 5 are what PreSonus Studio One calls logarithmic fades. The resulting crossfade is called an equal power crossfade. When crossfading two similar clips, the right fade type is probably closer to an equal power crossfade (Thornton n.d.).

Annica Milán says that sometimes the correction requests she receives are funny examples of the mind reading what it wants. Picture 8 presents one of her favourite correction requests she has received. It is a simple mix-up of one letter, but it makes the sentence very different. Milán says this is one of the funniest or most memorable corrections she has had to make.



Correction 12 ☑ Narrator □ Audio engineer

File nr.	Page nr.	Timestamp	Faulty word/sentence/sound
3	222	1:00:18	Mari (vahingossa uusi sivuhahmo)

Comment/instruction

Lukija: "Mari"
Teksti: "Meri"

Context

Meri lepäsi yhtä harmaana kuin taivaskin, täynnä jäälohkareita.

PICTURE 8. One of Annica Milán’s favourite correction requests (Milán 2022)

3.2 Pauses

A common practice is to have room-tone pauses at the beginning of a file and at the end of a file. Also, section breaks in the middle of the text have a pause to give a cue to the listener.

It is important that the narrator does not start narration at once when pressing record but leaves a pause of a few seconds at the beginning and at the end. It is also important that nothing happens in these pauses. There should be no breathing, moving around, writing on the keyboard, or anything. These clean room tones can be used to fill spaces that result when editing away extraneous sounds or lengthening a pause between the chapter title and the chapter start.

Audiobook Creation Exchange (n.d.) explains that these pauses at the beginning and at the end are to ensure that all files are successfully encoded into the different formats the book is delivered in. The second reason is that they too, like the section break pause, serve as an audio cue to the listener.

3.3 Extraneous noises

This category includes all kinds of random sounds. Clicks, bumps, stomps, loud breaths, barks, and meows. The workflow with these sounds is simple: the offending sound must be cut away from the narration and replaced with a clean room tone from somewhere else. Preferably the room tone comes from the same recording, so it sounds as similar as possible. Then the room tone is crossfaded with the rest of the narration, much like explained in chapter 3.1.

Of course, the above method is only possible if the offending sound happens outside of a sentence. If the sound happens during a sentence, the best course of action is to have the narrator retake the part. Trying to edit an offending sound out of the narration can take a long time and will usually end up with an obvious edit. It is much easier and faster to just do a retake.

3.4 Plosives

It is hard to stay still and keep the same distance to the microphone over a long period of time. Also, sometimes the narrator just uses more air for a certain word that has a plosive sound. The human voice is not a synthesiser, and variation is a part of its nature.

When proof listening a book, the editor might hear a burst of air that appears as a low-frequency bump. In editing and mastering, these are called plosives. In the Finnish language the unvoiced tenuis consonants p, t, and k are the main culprits for an obtrusive plosive. The reason is that these plosives are formed by stopping the airflow in the voice tract and then releasing it, which results in a burst of air (Savolainen 2001). A decent pop filter in front of the microphone will dampen these bursts of air before they reach the microphone capsule, but sometimes one or few get through.

The first plugin in the mastering chain, an equaliser set for high-pass filtering, will tame some of the milder plosives. But according to iZotope (n.d.), their De-plosive module, which is a part of the RX suite of plugins, works better when set before a high-pass filter. This is because the plugin tries to detect plosives in the range of 20 to 80 Hz, most of which is already cut by the high-pass filter, depending on the range of the voice.

Because how the De-plosive module tries to find the plosives, and because it is easy to set up too aggressive so that it affects healthy parts of the recording too, it is better to fix the most obtrusive plosives in the editing phase. The workflow in Studio One for fixing plosives is to cut the sentence that has the plosive and apply the RX De-plosive plugin as Event FX, so it applies to the separated sentence only.

The De-plosive plugin has three controls: sensitivity, strength, and frequency limit. The first control has the biggest effect on how effective the plugin is. The higher the sensitivity, the more it will detect plosives. But with higher sensitivity the plugin will also start to affect the overall quality of the signal, which is undesired.

iZotope (n.d.) says that higher values of the strength parameter can result in significant plosive reduction but can also negatively affect the rest of the signal. According to iZotope, it is better to use higher sensitivity with a lower strength.

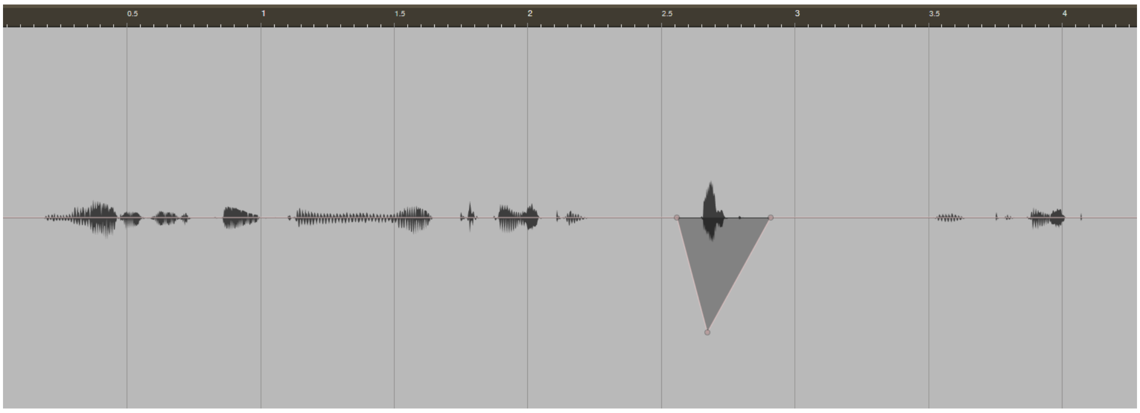
The last control, frequency limit, affects how high frequencies the plugin will operate on (iZotope n.d.). The default setting is 200 Hz, under which most plosives happen. Sometimes they appear lower or higher than that, so one needs to adjust accordingly. But the default setting is a good start.

If the problematic plosives cannot be fixed in post processing, then it is better to ask the narrator for a retake rather than damage the signal so that the edit sounds obvious. Because the listener might be very engrossed in the story, a weird edit could ruin the immersion.

3.5 Peaks

Sometimes there are words that are read much louder than the rest of the narration. An experienced narrator may notice these while recording and retake a part, but even experienced narrators miss these sometimes. The solution is to either cut and separate the loud part from the others and lower the gain or use a volume envelope.

Studio One has a feature called clip gain envelope, which makes it easy to control the gain of even a single word (Gilder 2020). It works by points and curves, basically like automation, but has the added benefit of visualised gain changes. Picture 9 shows the clip gain envelope in action.



PICTURE 9. The clip gain envelope in Studio One. A loud word has been brought down in volume, so it is more in line with the rest of the narration (Seppänen 2022)

A possible workflow with the clip gain envelope could be to add three points near the loud peak: one before the peak, one at peak, and one after it. The before and after points should be placed on silence/room tone, so the gain changes are limited to the peak only. Then the middle point is dragged down to a suitable level. The mid point’s placing may take a few tries to see where exactly the best spot for it is. Transparency is key, and the edit should not sound obvious.

The clip gain envelope can also be used if a single file has narration recorded with different levels, either due to changed preamp gain or microphone distance.

Then the workflow would be to place two points before the louder or quieter part. The first point ensures the level of the other part does not change, while the second will be used to control the level of the latter part. The same can be achieved by splitting the file in two and adjusting their gains, but the clip gain envelope is a useful tool if more precision is needed, and the editor would rather not split the file into too many parts. In Studio One, the clip gain envelope can also be disabled.

4 MASTERING

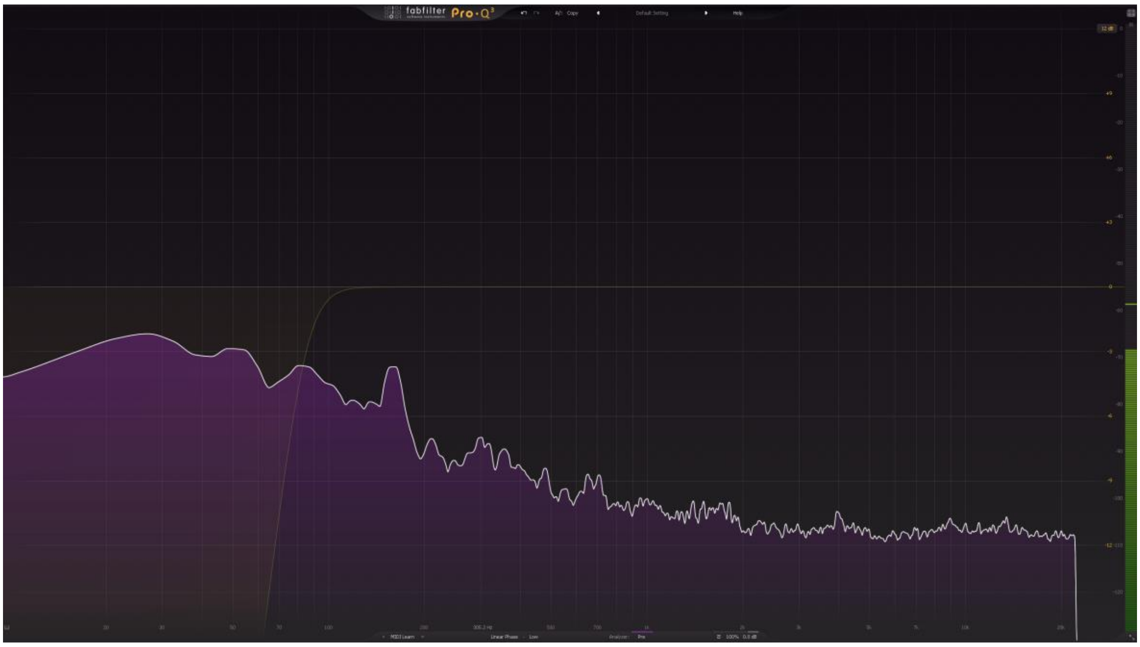
4.1 Equalizing or high-pass filtering

The first effect plugin in the processing chain is an equalizer, or EQ for short. The first EQ is used primarily as a high-pass filter. This means that the EQ removes frequencies below a set cut off frequency. High pass filtering is a fundamental principle in all mixing work (Senior 2011). When mixing music, almost every instrument, except for bass instruments, is high passed to some degree.

In narration, the cut off point for the filter is somewhere between 55 to 85 Hz. With a typical male voice, the point is closer to 55 Hz. And with a typical female voice, the frequency is closer to 85 Hz. The exact frequency depends on the range of the voice and where the fundamental frequency of the voice is.

The lowest frequencies are removed because they do not carry any useful information. In fact, they often hold data that is not only unnecessary but also undesired. For example, HVAC, traffic, and home appliances can cause noise in those lowest frequencies. Low frequencies travel farther, so a street outside the recording location can be a prominent source of low frequency noise (Dominic n.d.) Picture 4 shows a typical frequency distribution of noise in recordings. For this example, the cut off point for the filter has been set to 85 Hz. The filter is the faint, vertical greenish line on the left.

Picture 10 shows that there is a lot of information under 85 Hz, but virtually nothing that contributes to the narration in this book. The narrator of this book has a young female type of voice. According to Russel (2020), the average fundamental frequency for a female voice is 200 Hz, 125 Hz for a male, and 300 Hz for a child’s voice. These are not absolute values but merely ranges since speech is not a stable sine wave. Many things affect the fundamental frequency, such as age, hydration, and time of day. Sex is a rough categorisation by which clients often start to find a good match for a text.



PICTURE 10. Averaged frequency chart of background noise in a book. The chart presents frequencies from 0 Hz to 30 000 Hz (Seppänen 2022)

The analysis in Picture 7 is exaggerated. The analyser is set to show signal levels down to -120 dBFS, which makes the noise look loud. The actual level of the background noise in the narration is around -60 to -70 dBFS, which is a perfectly fine level and causes no problems with editing and mastering. The analyser's tilt setting is set to 0 dB per octave and white noise would appear almost ruler flat with these settings.

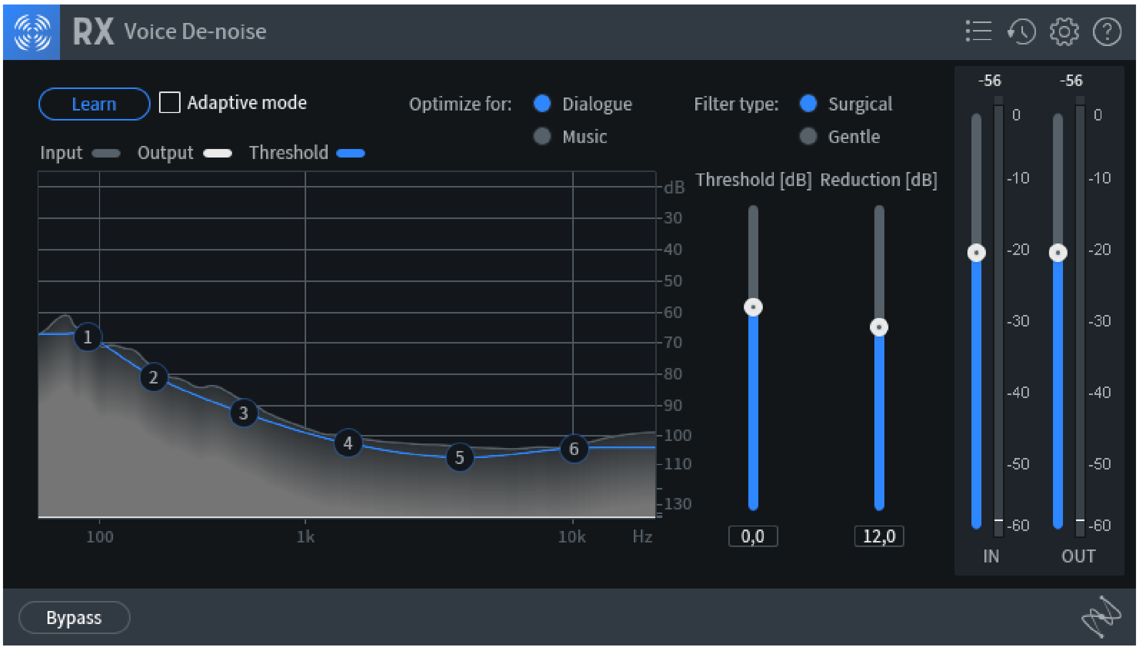
When using an equaliser in mastering and especially on low frequencies, it is advisable to use a linear phase EQ. Producer and composer Swisher (n.d.) explains that EQs cause a slight delay, and this leads to phase distortion. Phase distortion or smearing means that certain frequencies are delayed by different amounts than others. In narration where there is only a single mono channel, the phase distortion is probably not audible, but it can shift the phase of the waveform and cause problems with headroom later in mastering.

iZotope recommends using a high-pass filter before the next stage, noise reduction (n.d.). Since with the high pass filter some obvious, offending low frequency rumble has been removed, the noise reduction plugin has a cleaner signal to focus on.

4.2 Noise reduction

A common noise floor target for audiobooks is -60 dB maximum in the finished file. Some companies request an even lower noise floor, but ACX (n.d.) and NNELS or The Canadian National Network for Equitable Library Service (n.d) request the maximum of -60 dBs.

The best place to treat the noise is before recording, as described in chapter 2.1.1. But getting a very low noise floor with a room that was not specifically built for that can be difficult. If the recorded noise floor does not quite reach as low as -60 dB, or whatever the client requests, a noise reduction plugin can be used. iZotope's RX Voice De-noise shown in Picture 11 is a common tool for noise reduction in dialogue editing but also narration editing.



RX Voice De-noise

Learn □ Adaptive mode

Optimize for: ● Dialogue ○ Music

Filter type: ● Surgical ○ Gentle

Threshold [dB] 0,0

Reduction [dB] 12,0

Bypass

PICTURE 11. iZotope RX Voice De-noise with a room tone running through it (Seppänen 2022)

The workflow with Voice De-noise is to loop a section of the narration where there is absolutely nothing happening but just the noise floor. In most practical situations, the noise floor level is entirely dictated by the environmental noise: HVAC, home appliances, and outside noise – to name a few common culprits.

iZotope (n.d.) says that Voice De-noise consists of 64 bandpass filters which act as a multiband gate to pass or stop a signal based on the nodes' threshold values. If a signal component falls under a specific level, it will be attenuated. If a signal component goes above the threshold, nothing will be done to it. This process is also sometimes referred to as gating or downward expansion.

When the noise is looping, one needs to press the Learn-button and let the plugin study the noise for a few seconds. When that is done, Learn needs to be unpressed. (iZotope n.d.)

With Voice De-noise, the maximum gain reduction is 20 dB, which should be enough if most of the noise was taken care of before the recording phase. If 20 dB does not seem to be enough, a second Voice De-noise can be tried in series. But there is a chance that aggressive noise reduction starts to adversely affect the narration, especially the ends and beginnings of words. This is one of the reasons why a narrator must carry all sentences with a good energy to their end, so that noise reduction can work properly.

4.3 De-essing

Particularly condenser microphones with their sensitivity to high frequencies tend to pick sibilance very well. Sibilance or hissing is often referred to as a harsh, piercing quality of audio. Sibilance is a quality of speech that sticks out and can become burdensome to listen to if left unprocessed. Therefore, a de-esser is used nearly every time in mastering an audiobook.

In Finnish phonetics, sibilant sounds belong to the category of fricatives. The category of fricatives includes f, s, š and h. (Savolainen 2001.) S and š are probably the most obvious of these. A good starting point for setting up a de-esser is to find a sentence of the narration that has a lot of s-sounds. The Finnish tongue twister “Vesihiisi sihisi hississä” is a prime example of sibilant sounds.

According to Susic (2020), sibilance happens between 5 and 8 kHz. It can also appear below or above that range. Male sibilance also tends to happen a bit lower than female sibilance. But between 5 and 8 kHz is where one should start looking for the offending frequencies. (Susic 2020.)

In music production, it is common to make precise gain automation for sibilant sounds. But audiobooks are generally hours in length and such detailed work would take atrocious amounts of time. Therefore, a de-esser needs to be used.

De-essers can be split into wide band and split band. Messite (2021) explains that a wide band de-esser turns down the entire signal when it detects sibilance, while a split band de-esser splits the signal into different frequency ranges and turns down the gain of a range when the signal exceeds the set threshold in that range.

According to Kaul (n.d.), wide band de-essing can sound more natural, since it affects all frequencies equally. But it is also harder to set for specific frequencies. Kaul (n.d.) also says that split band de-essers can be set for more surgical processing, since many modern de-essers allow to change the frequency range where the plugin operates.

De-essing can be thought of as a form of dynamic EQ or multiband compressor: all three methods lower the gain of either the whole signal or a specific range once the signal exceeds a given threshold in a specified frequency range. The De-ess module in iZotope’s RX is a common piece of software for de-essing. It has both a wide band and a split band setting. iZotope's De-ess module calls wide band operation “Classic” and split band mode “Spectral” (iZotope, n.d.).

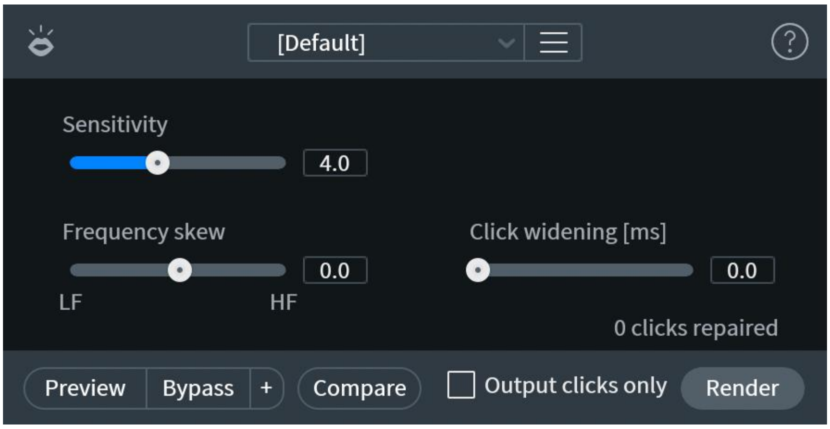
Messite (2021) recommends against using only one de-esser with aggressive settings. According to Messite (2021) it is better to do it in series with each de-esser taking away a couple of dBs at most. If the first EQ whose primary function is being a high-pass filter can also do dynamic equalizing, one can apply mild de-essing already with that one. FabFilter’s Pro-Q 3 is an EQ that can do both dynamic and static equalizing. It is worth noting that static equalizing, which means that a cut or boost is constantly applied to a specific range of frequencies, does not work as a de-esser. A static boost or cut applies to all sounds in the defined range, not just the offensive sibilants, and thus easily makes the narration sound unnatural.

4.4 Mouth clicks

Many people listen to audiobooks on their headphones. Therefore, it makes sense to complete the mastering on headphones, so it translates better to all the different models of headphones and earphones listeners use. Headphones can bring attention to problems that cannot be heard as clearly with monitor speakers (Guttenberg 2016).

A condition called misophonia exists. WebMD described the condition as “a strong dislike or hatred of specific sounds.” (2020). People with the condition often say that they are triggered by oral sounds, such as mouth clicks. And mouth clicks are a thing every narrator has to some degree. To make audiobooks more accessible, a plugin that reduces or removes mouth clicks can be used. And even if the listener does not have misophonia, excessive mouth clicks take away from the listening experience anyway.

iZotope’s RX suite of plugins has a module called Mouth De-click. One or two instances of the plugin on the narration can reduce or even virtually remove mouth clicks and make for a more enjoyable listening experience. Picture 12 shows the interface of the plugin. The plugin has only three adjustable settings.



Sensitivity 4.0

Frequency skew 0.0

Click widening [ms] 0.0

0 clicks repaired

PICTURE 12. iZotope RX Mouth De-click -plugin (iZotope n.d.)

The three settings are sensitivity, frequency skew and click widening. iZotope (n.d.) says that sensitivity affects how many clicks are detected. Frequency skew affects what kind of clicks are targeted, and according to iZotope (n.d.), settings of zero and above target mouth clicks in the middle frequencies. Click widening adjusts the repaired are around the clicks, since some mouth clicks are longer than others (iZotope n.d.).

4.5 Second EQ

A second equalizer is most of the time not necessary. However, depending on which microphone was used and how the de-essing is set up, the voice may lose some high frequencies.

As shown in Figure 5 on page 36, the frequencies most important for speech intelligibility in Western languages are in the range of 1 to 4 kHz. Conversely, frequencies above or below this range are not imperative for understanding. If the voice was band limited to have nothing but the frequencies 1 to 4 kHz, it would probably be understandable. But without those “supporting” frequencies, the voice would not sound natural. Band limiting a signal to around those frequencies is a sound design technique to make something sound like it is coming from a radio or a small speaker (iZotope 2016).

In addition to de-essing dulling the high frequencies, if the narration was recorded with a dynamic microphone, such as the Shure SM 7 B, a high frequency boost may be worth considering. Dynamic microphones are less sensitive to higher frequencies, as studied in chapter 2.3.2. Something like a static high-shelf with a maximum of +3 dBFS from 5 kHz should show if that area needs boosting at all. It depends greatly on the voice. A high-shelf boost is a common thing in music production, but in an audiobook the voice is not fighting for its space with anything else. Boosting is something that is worth a quick try, but if it does not seem to make the voice sound better, it is probably best not to do it.

The recording may have overemphasised low frequencies too. This sometimes happens when someone with a very low voice range speaks close to the microphone and either consciously or unconsciously abuses the proximity effect of the microphone. Booming low frequencies are not equally as annoying as piercing high frequencies, but equalizing is all about finding the balance. If the voice appears to be bass-heavy, a static low-shelf EQ with a maximum of -3 dBFS setting at around 200 Hz is a good place to start.

It is worth noting that static equalizing a wide range of frequencies is a powerful tool and can make something sound off fast. If it appears that a recording needs something like -6 dBs or +6 dBs (Full Scale) in a certain frequency range, something in the recording process is probably not right. The first place to start looking for such problems is microphone placement relative to both the room and the mouth.

4.6 Compression

Compression, if necessary, should be applied mildly. Managing Director of audiobook production house Ladbroke Audio, Neil Gardner, says that in audiobook mastering “less is more, allow for dynamic range” (2016). Gardner also advises subtlety in EQ, compression and limiting (2016). If one were to apply compression, it should be as transparent as possible. Post-processing narration should never sound obvious. Very high peaks should be turned down in the editing already. But if the narration is very dynamic, then this becomes unfeasible.

The narrator should be able to deliver a performance that is consistent in volume. The volume level may slightly change from chapter to chapter, but the level should not fluctuate too much inside a single chapter. It is easy to lower or increase the gain of a single chapter in the mastering process, so all the files are in the same ballpark what it comes to loudness.

If the narration has a lot of volume fluctuation, a compressor with mild settings can be used. Mild settings could be something like a maximum ratio of 2:1 with a soft knee. The threshold should be set so that most of the time the compressor is not active at all. An aggressive compressor can also bring up the noise floor, which was piously lowered in the second stage of mastering. Audiobook production, unlike music production often, is not a race to achieve maximum loudness.

4.7 Limiting

On the Audiobook Creation Exchange blog, marketing and communications manager Jacobi (2014) advises that instead of a compressor, a limiter should be used to achieve the desired loudness level. Jacobi's reasoning is that while a compressor can achieve similar things, it is easier to use improperly.

ACX's (n.d). submission requirements list the following on loudness:

- Maximum peak value of -3 dB
- RMS level between -23 dB and -18 dB.

While ACX talks about RMS level, LUFS is another common metric. iZotope RX Audio Editor can batch-process audio with a custom module chain. In the Loudness Control -module one can set the desired true peak, integrated LKFS, which today is the same as LUFS, and tolerance, which means the margin of error for loudness units.
In the batch processor, one also must specify the output bit depth and dither settings.

With the desired settings in place, RX will either lower or increase the gain of each file separately, so they meet the required levels. If a true-peak level has been specified, a post-limiter will be applied to the signal to meet the specification (iZotope n.d.). Adjusting each file to be equally loud is important so that the listener does not have to adjust their volume all the time but can trust that the volume will be roughly the same through the book.

In this last stage, dithering should also be applied if the files are going to be exported in a lower bit depth. The bit depth is also selectable in RX's batch processor. RX dithers the files in the same process as the loudness control, making the process simple. Generally, the final delivery format will be 44.1 kHz sampling rate with 16-bit depth. So, if the recording was done in 24-bit depth and the final export is going to be 16-bit, dithering should be applied.

When RX has run its batch processor, one should take a final listen to the files and possibly run them through an analyser to make sure every file is equally loud or at least within tolerance. In the final listening, it is important to look at the noise floor and make sure it has not risen in volume due to the gain adjustments done in limiting. The highest peaks are also important to check so that there is not any audible distortion. Neil Gardner (2016) says to listen for distortion, compression artefacts and the frequency balance in the final product, but also after every step of the process.

This stage of mastering is the final touch. As mentioned before, recording is where most of the work is done and that is the stage where everything needs to be right. In narration there is no possibility to hide ugly processing behind a wall of instruments or a wash of reverb. The recording of the voice needs to be able to stand on its own.

4.8 Delivery

The final delivery format will depend on the client, but 44.1 kHz 16-bit depth WAV is one standard. It is a standard for music, and it is also the standard for a variety of audio works. Lossless WAV with a true peak of at least -1 dB can safely be converted into lossy formats, such as 192 kbps MP3. Audiobook streaming services often use lossy compression. (BookBeat 2022.)

Whatever the format requested by the client is, it is a good practice to save and archive the original recordings and the lossless master. This way a new master can be made if the first master does not meet the requirements, the files are corrupted, or if anything unexpected happens to it. The files should also be backed up in multiple places. An adage says that if the files are not in three places, they do not exist (Alison 2013).

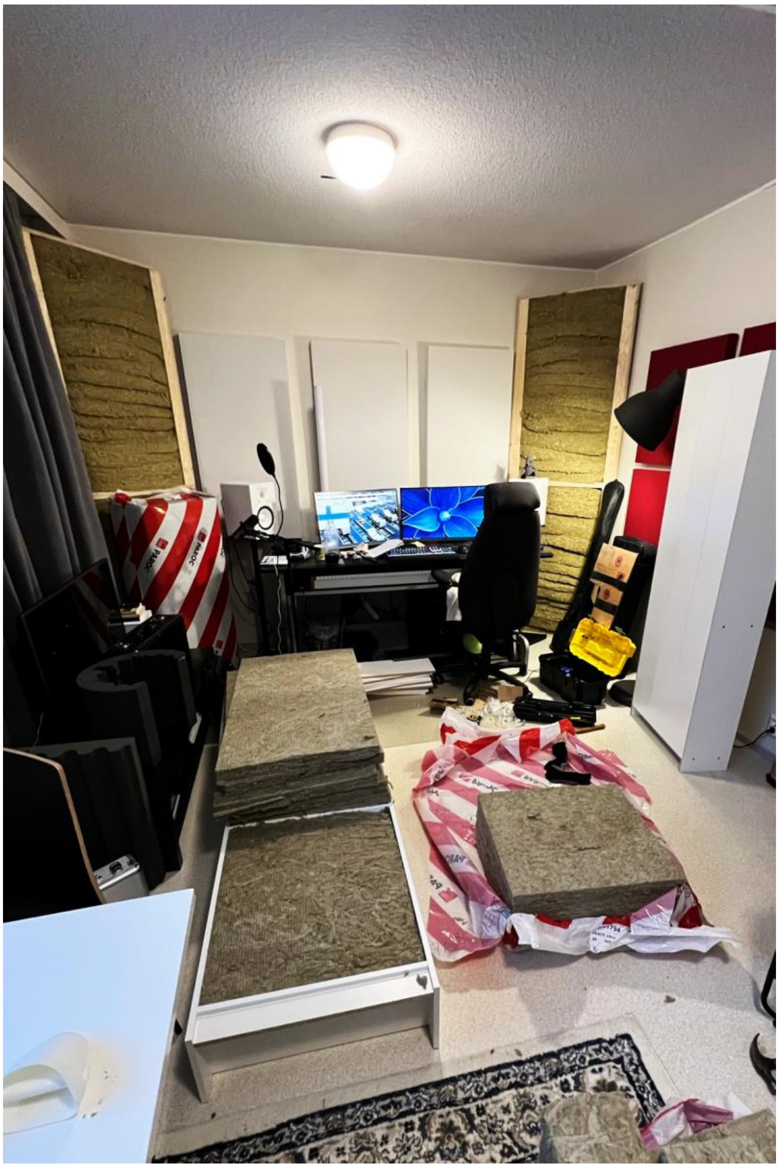
In 2022, files can be shared through a cloud service. Examples of cloud services are Google Drive, Microsoft OneDrive, and WeTransfer. Some companies may also have their own servers and require uploading via FTP. FileZilla is one free, open-source FTP software.

5 RECORDING THE BOOK “OUTO INTOHIMO”

I recorded this title in my home, which is a studio apartment in an apartment building built in the 60’s. Even though it has problems with sound proofing, it does not have the problem of other people barging into the studio space when working, which can be a problem with multi-user spaces. At home I can work in relative peace and at my own pace.

Outo intohimo was the 49th title I recorded. Most of my earlier titles were recorded with the Aston Spirit, which is an LDC microphone. But due to LDC microphones’ higher sensitivity to ambient noise together with my apartment’s problems with sound proofing, I switched over to the Sennheiser MKH 416, which is a shotgun microphone. The MKH 416 rejects off-axis noise better thanks to its tight super-cardioid polar pattern, as studied in chapter 2.13. If I were recording in a space that is very well treated, I would use almost any LDC microphone, but in my home, I cannot have all the walls treated, so I must compromise.

Most of my recording space is treated, though. Picture 13 from a building session shows the acoustic treatment I have. In the corners there are bass traps, that have since been covered with fabric. The fabric to the left is a heavy curtain that is commonly used in theatres, for example. On the wall behind the screens there are broadband absorbers. The red tiles on the right are also broadband absorbers. The things on the floor are gobos built from IKEA shelves and mineral wool.



PICTURE 13. Building bass traps and gobos for a home studio (Seppänen 2022)

The audio interface I used for this title and all other titles is the PreSonus 1824c. Considering the channel-count, it is overkill for a project like this, but I use it for everything else too and it is always on my desktop. I also have a Focusrite Scarlett 2i2, which I use if I must record in another location.

The book is by E. T. A. Hoffman, who lived in 1776-1822. The story happens in Paris during the reign of Ludwig XIV and tells of a wave of robberies happening in the city. The e-book version is eighty-three pages long and the audiobook format is about 3 hours and 9 minutes long. (BookBeat n.d.)

Probably the most challenging aspect of this book were the French names and their pronunciation. In narration I aim to keep the same pronunciation as in the original language as closely as possible. Thus, I contacted a French-speaking friend, who aided me over the internet in achieving a close-enough pronunciation.

I split the narration into multiple sessions, aiming to have at least 30 minutes of newly recorded material after every session. I split the sessions like this not because of my voice tiring, but due to time constraints and a busy schedule. Due to the sound-proofing issues, I also must record mostly at night, which affects the length of my sessions somewhat.

When I start recording, I block the loudest air vents of my apartment with a potholder and a piece of neoprene foam. The foam works with the return vents well because it is light enough to stay sucked in but dense enough to really block the sound. I have also built a system for turning off the fridge with Philips Hue. The fridge gets its power via a Philips Hue smart plug, which I control via Bluetooth on my phone. This means that I do not have to climb a chair every time to unplug the fridge. The effect on background noise with these precautions can be seen in Picture 2 on page 12.

I have both the text and my recording software, Studio One, on one 27-inch monitor. The text takes up the left half of the monitor and software the right half. I have two screens, and both of my screens are off axis, but the microphone is right on the centre. This way the microphone does not block the screen. For a pop filter I use the windscreen that came with the Sennheiser MKH 416. Even though it may dampen the highs a bit, it is useful because it does not block my view at all.

Different studios may have different requirements, but the one I deliver to requires having a three second pause between chapters. I have setup a macro in Studio One for this. I place the playhead at the end of the last word in a chapter, click the macro and it automatically adds a marker three seconds from the playhead. The same macro also moves the playhead forward half a second, so I can have a clean crossfade right in the middle of the room noise. I have similar macros for other lengths, which I use when editing books narrated by others.

I aim to have one file be a maximum of 120 minutes in length. In my recording template I have a marker at this spot, so it is easier to not go over the time. If I finish a chapter somewhere around 105 minutes to 120 minutes, I start a new project file. If I go over the 120-minute limit, I will export the last chapter separately and add it to the beginning of the next file.

After I have finished narrating the book, I go through every project file to look for unnecessary sounds, such as obvious mouth clicks or transient noises. I also use another macro that lengthens all crossfades and makes them a bit equal powered. Then I manually check all the crossfades to see that all of them happen in room tone and nothing overlaps the narration.

Then I export the files in 44.1 kHz and 16-bit WAV-format. I upload them to Google Drive and send the link over email to the engineer. I do not engineer the books I narrate, except for the mandatory edits with the length of the pauses and crossfades.

The finished books are published on streaming services usually within a few weeks of me delivering the files to the engineer. The files for Outo intohimo I delivered on the 22nd of March 2022, and the finished audiobook was available on BookBeat and other platforms on the 13th of April 2022.

6 DISCUSSION

The focus of this thesis was to study the different aspects of audiobook production and their role and importance in the process. Especially the recording stage was given attention because that is where things must be done right for the process to succeed. Since many narrators are recording from home by themselves, there are many things that must be understood by a single person.

The study confirms that most of the work is done in the recording phase. Of utmost importance is finding a suitable recording space: a room that is not too reverberant or has a high noise floor. Building a room within a room is expensive and outside many budgets, but even a walk-in closet can be turned into a vocal booth in which professional quality recordings can be made, as proven by Annica Milán's setup. A voice sample for an audition can be recorded even on a cruise ship (Milán 2022).

When putting together a recording setup, most of the budget should go towards acoustic treatment and the microphone. But even the microphone does not need to be a top-of-the-line Neumann: Milán recorded her first books with a Rode NT1-A and recorded her voice sample on a cruise ship with friend's mobile equipment (Milán 2022).

The study also confirmed that if recording is done properly, editing and mastering are mere final touches. The purpose of these latter stages is to ensure there are no mistakes in the narration, clean up the audio, maybe enhance it a bit with EQ and compression, and bring all the files to an equal loudness so the listener can be engrossed in the story.

Technology does not make one a narrator, but the ability to deliver text in a neutral yet interesting way does. One way to look at the reasoning of an expensive microphone or preamp is that with money a cleaner sound can be achieved, and the narrator's voice can be presented as it is without post-processing adversely affecting it. But there is no reason one could not start with a cheaper setup. The recording environment is more of a limiting factor than the price of the microphone.

Most of the findings of this study are directly applicable to podcast-recording and other types of voice-overs, even if not explicitly said in the text. Stylistic aspects and matters related to voice-over direction were intentionally left out of this thesis, as they are a matter that would deserve their own theses. The focus of this thesis were the technical aspects of audiobook production.

REFERENCES

AccessiblePublishing.ca. N.d. Audiobook Recommendations for Publishers. Read on 26.02.2022. https://www.accessiblepublishing.ca/audiobook-recommendations-for-publishers/

ACX. N.d. ACX Audio Submission Requirements. Read on 25.02.2022. https://www.acx.com/help/acx-audio-submission-requirements/201456300

Alison. 2013. Ballyhoo. If it’s Not in Three Places it Doesn’t Exist. Released on 22.01.2013. Read on 25.02.2022. https://ballyhoo.co.uk/If-its-Not-in-Three-Places-it-Doesnt-Exist/

Archtoolbox. 2021. Architectural Acoustics – Acceptable Room Sound Levels. Updated on 12.05.2021. Read on 13.02.2022. https://www.archtoolbox.com/materials-systems/architectural-concepts/architectural-acoustics-acceptable-room-levels.html

Audacity. N.d. Audacity Manual. Fade and Crossfade. Released on 16.11.2021. Read on 19.02.2022. https://manual.audacityteam.org/man/fade_and_crossfade.html

Audacity Wiki. 2019. Hardware influence on recording quality. Read on 14.04.2022. https://wiki.audacityteam.org/wiki/Hardware_influence_on_recording_quality

Audio Technica. N.d. Audio Solutions Question of the Week: What Is the Difference Between a Windscreen and a Pop Filter? Read on 12.04.2022. https://www.audio-technica.com/en-us/support/audio-solutions-question-of-the-week-what-is-the-difference-between-a-windscreen-and-a-pop-filter/

Bashforth, E. 2021. Patient. Why does your stomach rumble when you aren’t hungry? Reviewed by Dr Sarah Jarvis MBE. https://patient.info/news-and-features/why-does-your-stomach-rumble-when-you-arent-hungry

Becker, M. 2016. Vetstreet. Cat Always meows to Go Out? 6 Ways to Stop It. Read on 12.04.2022. https://www.cathealth.com/cat-care/training/2525-how-to-teach-your-cat-not-to-meow-to-go-outside

BookBeat. N.d. Kaikkien aikojen suosikit. Read on 15.02.2022. https://www.bookbeat.fi/kirjalista/kaikkien-aikojen-suosikit-94760

BookBeat. N.d. Lukijat: Annica Milán. Read on 25.02.2022. https://www.bookbeat.fi/lukijat/Annica%20Mil%C3%A1n

BookBeat. 2022. Email. 04.02.2022.

BookBeat. Outo intohimo. Read on 12.04.2022. https://www.bookbeat.fi/kirja/outo-intohimo-646566

Boudreau, M. The Podcast Host. 2020. Mic Technique for Podcasters | How to Sound Your Best. Released on 21.01.2020. Read on 19.02.2022. https://www.thepodcasthost.com/recording-skills/mic-technique-for-podcasters/

Burke, S. GamersNexus. 2011. The Basics of Case Fan Noise, Airflow, and Quieter gaming. Released on 30.12.2011. Read on 18.02.2022. https://www.gamersnexus.net/guides/695-basics-of-case-fan-noise-and-airflow-quieter-gaming

Ciccarelli, S. 2018. Voices.com Microphone Setups – How To Find The Sweet Spot. Read on 14.04.2022. https://www.voices.com/blog/microphone_sweet_spot/

Ciccarelli, S. 2020. Voices.com. How To Pick the Right Microphone For Your Voice. Read on 14.04.2022. https://www.voices.com/blog/how_to_pick_the_right_microphone/

Clark, N. Voices. 2018. Styles of Voice Over Narration. Released on 07.06.2018. Read on 16.02.2022. https://www.voices.com/blog/styles-of-voice-over-narration/

Clement, S. LIVESTRONG. N.d. How Many Calories Do I Lose by Talking? Read on 17.02.2022. https://www.livestrong.com/article/319364-how-many-calories-do-i-lose-by-talking/

Computer Lounge. 2019. Ultimate Keyboard Showdown: Mechanical vs. Membrane Keyboards. Read on 14.04.2022. https://www.computerlounge.co.nz/blog/tips-and-tricks/ultimate-keyboard-showdown-mechanical-vs-membrane-keyboards

Crucial. 2017. Why is my Computer Loud? Read on 12.04.2022. https://www.crucial.com/articles/pc-builders/why-your-computer-is-loud-and-how-to-reduce-noise

DelGaudio, M. 2017. YouTube video. $1000 Mic Shootout - Sennheiser MKH416 vs Neumann TLM 103. Published on 14.01.2017. Referred on 14.04.2022. https://www.youtube.com/watch?v=sdE_VekATvE

Dominic. N.d. Soundproof Central. How To Block Low Frequency Sound Waves (Bass). Read on 12.04.2022. https://soundproofcentral.com/block-low-frequency-sound-waves/

Doroudiani, S. 2015. Quora. Why are synthetic fabrics often louder than natural ones? Read on 14.04.2022. https://www.quora.com/Why-are-synthetic-fabrics-often-louder-than-natural-ones

DPA Microphones. 2019. The Basics about Comb Filtering (And How to Avoid It). Read on 12.04.2022. https://www.dpamicrophones.com/mic-university/the-basics-about-comb-filtering-and-how-to-avoid-it

DPA Microphones. 2021. Facts About Speech Intelligibility. Released on 03.03.2021. Read on 10.02.2022. https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility

Enoch, A. HouseClap. 2019. How Can I Stop My Fluorescent Lights From Buzzing? Released on 02.08.2019. Read on 13.02.2022. https://www.houseclap.com/how-can-i-stop-my-fluorescent-lights-from-buzzing/

Focusrite. N.d. Focusrite Scarlett 2i2 User Guide. Read on 08.02.2022. https://resource.focusrite.com/sites/default/files/focusrite/downloads/7317/scarlett-2i2-user-guide-v2.pdf

Fox, A. N.d. What Is Phantom Power And How Does It Work With Microphones? Read on 08.02.2022. https://mynewmicrophone.com/what-is-phantom-power-and-how-does-it-work-with-microphones/

French, N. & Steinberg J. 1947. The Journal of the Acoustical Society of America 19 (1). Read on 10.02.2022. https://jontalle.web.engr.illinois.edu/uploads/537.F18/Papers/FrenchSteinberg47.pdf

Fuston, L. Sweetwater. 2021. Released on 03.05.2021. Read on 18.02.2022. https://www.sweetwater.com/insync/what-is-it-comb-filtering/

Gardner, N. 2016. Ladbroke Audio. The Art of Audiobook Mastering. Read on 14.04.2022. http://www.ladbrokeaudio.com/the-art-of-audiobook-mastering/

George, E., Festen, J. & Houtgast, T. 2008. The Journal of the Acoustical Society of America. Read on 12.04.2022. The combined effects of reverberation and non-stationary noise on sentence intelligibility. https://asa.scitation.org/doi/pdf/10.1121/1.2945153

Gilder, J. 2020. YouTube video. How Clip Gain Envelopes Work in #StudioOne. Published on 21.07.2020. Referred on 14.04.2022. https://www.youtube.com/watch?v=WZVbb7Uz9gI

Grace Design. N.d. Grace Design M101 Single Channel Microphone Preamplifier Owner's Manual Rev A. Read on 08.02.2022. https://www.gracedesign.com/support/m101_manual_RevA.pdf

Guttenberg, S. 2016. CNET. What's more accurate: Speakers or headphones? Read on 12.04.2022. https://www.cnet.com/tech/mobile/whats-more-accurate-speakers-or-headphones/

Harding, S. 2019. Tom's Hardware. Logitech MX Master 3 Wireless Mouse Review: Reinventing the Wheel Successfully. Read on 14.04.2022. https://www.tomshardware.com/reviews/logitech-mx-master-3-wireless-mouse,6311.html

Harman. 2021. Last modified on 11.05.2021. Read on 12.04.2022. https://help.harmanpro.com/what-is-a-pop-filter-for

Harris-MacDuff, A. 2021. Voquent. Best Interfaces & Pre-amps for Voice-Over. Read on 14.04.2022. https://www.voquent.com/best-interfaces-pre-amps-for-voice-over/

Henshall, M. Shure. 2014. What Is Phantom Power & Why Do I Need It? Released on 28.02.2014. Read on 08.02.2022. https://www.shure.com/en-US/performance-production/louder/what-is-phantom-power-why-do-i-need-it

iZotope. N.d. iZotope RX Audio Editor Manual. Read on 25.02.2022.

iZotope. 2016. Using Vinyl for an old radio sound. Read on 14.04.2022. https://www.izotope.com/en/learn/using-vinyl-for-an-old-radio-sound.html

Jacobi, Scott. 2014. Audiobook Creation Exchange (ACX). Released on 11.07.2014. Read on 25.02.2022. https://blog.acx.com/2014/07/11/how-to-succeed-at-audiobook-production-part-3/

J'vlyn d'Ark. 2017. LANDR Blog. What Is Reverb? The 8 Step Guide to Mixing's Most Powerful Effect. Read on 14.04.2022. https://blog.landr.com/what-is-reverb/

Katz, L. 2020. SoundGuys. Why conference calls sound bad. Read on 14.04.2022. https://www.soundguys.com/why-conference-calls-sound-bad-23723/

Kaul, Vinny. N.d. Producer Hive. The Definitive Guide to Working With De-essers (W/ Examples). Read on 24.02.2022. https://producerhive.com/music-production-recording-tips/what-is-a-de-esser/

Knuuttila, M. 2021. Iltalehti. Jos kärsit äänistä, vältä näitä asuntoja. Read on 14.04.2022. https://www.iltalehti.fi/asumisartikkelit/a/f0ceb84a-36b0-486f-b081-293f790e2e95

Koschak, M. Shure. N.d. Choosing a Shotgun Microphone: The Long and Short of It. Read on 10.02.2022. https://www.shure.com/en-US/performance-production/louder/choosing-a-shotgun-microphone-the-long-and-short-of-it

K.T., T. Lewitt. 2016. Microphone polar patterns. Released on 16.08.2016. Read on 10.02.2022. https://www.lewitt-audio.com/blog/polar-patterns

Levine, Z. 2014. Presentation Products. A Comparison of Conference Room Microphone Solutions. Read on 12.04.2022. https://www.presentation-products.com/a-comparison-of-conference-room-microphone-solutions/

Listening Books. 2019. What is proof listening? Our audiobook producer reveals all the secrets! Read on 14.04.2022. https://www.listening-books.org.uk/what-is-proof-listening

Masterson, K. National Public Radio. 2011. From Grunting To Gabbing: Why Humans Can Talk. Released on 11.08.2011. Read on 15.02.2022. https://www.npr.org/templates/story/story.php?storyId=129083762&t=1644950329531

Merlot, J. N.d. Reboot Recording. Find The Right Distance between Mouth and Mic Fast. Read on 10.02.2022. https://rebootrecording.com/distance-mouth-mic/

Messite, N. iZotope. 2021. What Is De-essing? The Dos and Don't's of Using a De-esser. Released on 10.12.2021. Read on 24.02.2022. https://www.izotope.com/en/learn/the-dos-and-donts-of-de-essing.html

Michael, R. N.d. FilmSound.org. Avoiding Mouth Noise / Mouth clicks. Read on 14.04.2022. http://filmsound.org/QA/mouthclick.htm

Milán, A. Singer-narrator. 2022. Interview on 22.02.2022. Interviewer Seppänen, J. Tampere.

Milán, A. Singer-narrator. 2022. WhatsApp-message. Sent on 22nd and 23rd of February 2022.

Monk, S. 2012. musicianself. How to make good vocal recordings? Mic positioning. Released on 17.08.2012. Read on 19.02.2022. https://musicianself.com/vocal-recording-and-mic-positioning/

Neumann. N.d. Microphone Data (3). What is Self-Noise (or Equivalent Noise Level?). Read on 13.02.2022. https://www.neumann.com/homestudio/en/what-is-self-noise-or-equivalent-noise-level

Neumann. N.d. Microphone Basics (5). What is the Proximity Effect? Read on 10.02.2022. https://www.neumann.com/homestudio/en/what-is-the-proximity-effect

Nursing Times. 2018. Every breath you take: the process of breathing explained. Read on 12.04.2022. https://www.nursingtimes.net/clinical-archive/respiratory-clinical-archive/every-breath-you-take-the-process-of-breathing-explained-08-01-2018/

Osburn, W. 2020. The Seasoned Podcaster. Best Microphone for an Untreated, Echoey Room. https://www.theseasonedpodcaster.com/gear/best-microphone-for-an-untreated-echo-room/

Peter. 2019. Microphone Geeks. Understanding different microphone polar patterns. Read on 12.04.2022. https://microphonegeeks.com/different-microphone-polar-patterns/

Robson, D. 2018. BBC. The reasons why women's voices are deeper today. Released on 13.06.2018. Read on 27.02.2022. https://www.bbc.com/worklife/article/20180612-the-reasons-why-womens-voices-are-deeper-today

Ramm, R. N.d. Orpheus Audio Academy. Dynamic vs. Condenser Mic: Which Is Better For Vocals? Read on 14.04.2022. https://www.orpheusaudioacademy.com/dynamic-vs-condenser-mic/

Royer Labs. N.d. Read on 12.04.2022. Ribbon Mics and Phantom Power. https://royerlabs.com/ribbon-mics-and-phantom-power/

Russel, J. 2020. Accusonus. The Human Voice and the Frequency Range. Released on 23.09.2020. Read on 23.02.2022. https://blog.accusonus.com/pro-audio-production/human-voice-frequency-range/

Samula, T. N.d. Realia isännöinti. Taloyhtiön lämmitys – usein kysytyt kysymykset. Read on 14.04.2022. https://www.realiaisannointi.fi/ajankohtaista/taloyhtion%E2%80%93lammitys

Savolainen, E. 2001. Finn Lectura. Verkkokielioppi. Klusiilit p, t, k, d, (b, g). Read on 12.04.2022. https://fl.finnlectura.fi/verkkosuomi/Fonologia/sivu151.htm

Savolainen, E. 2001. Finn Lectura. Verkkokielioppi. Frikatiivit s, h, (f, š). Read on 14.04.2022. https://fl.finnlectura.fi/verkkosuomi/Fonologia/sivu153.htm

Senior, M. 2009. Sound On Sound. Q. How much difference does mic position make to vocals? Released in April 2009. Read on 19.02.2022. https://www.soundonsound.com/sound-advice/q-how-much-difference-does-mic-position-make-vocals

Senior, M. 2011. Mixing Secrets for the Small Studio. Burlington, MA: Focal Press.

Sennheiser. N.d. MKH 416-P48U3. https://en-fi.sennheiser.com/short-shotgun-tube-microphone-camera-films-mkh-416-p48u3

Shure. N.d. Shure SM58 Vocal Microphone Guide. Read on 08.02.2022. https://pubs.shure.com/guide/SM58/en-US

Shure. 2021. SM7B Output Level and Preamp Gain Specifications. Updated on 07.06.2021. Read on 08.02.2022. https://service.shure.com/Service/s/article/sm7-output-level-and-preamp-gain-specifications?language=en_US

Stone, D. N.d. Laptop Vs. PC Power Consumption. Read on 12.04.2022. https://smallbusiness.chron.com/laptop-vs-pc-power-consumption-79347.html

Summers & Zim's, Inc. N.d. Supply vs Return Vents. Read on 13.02.2022. https://sumzim.com/supply-vs-return-vents/

Susic, P. 2020. HeadphonesAddict. What is Sibilance? Released on 02.09.2020. Read on 24.02.2022. https://headphonesaddict.com/sibilance/

Sweetwater. N.d. Grace Design M101 Half-rack Microphone Preamp. Read on 08.02.2022. https://www.sweetwater.com/store/detail/m101--grace-design-m101

Swisher, D. N.d. Musician on a Mission. Linear Phase EQ: The Dos and Don'ts of Linear EQ. Read on 23.02.2022. https://www.musicianonamission.com/linear-phase-eq/

Teach Me Audio. 2020. Condenser Microphone. Updated on 26.04.2020. Read on 08.02.2022. https://www.teachmeaudio.com/recording/microphones/condenser-microphone

Teach Me Audio. 2020. Dynamic Microphone. Updated on 02.05.2020. Read on 08.02.2022. https://www.teachmeaudio.com/recording/microphones/dynamic-microphone

Thomann. N.d. AKG C414 XLS. Read on 08.02.2022. https://www.thomann.de/fi/akg_c414_xls.htm

Thomann. N.d. Audio-Technica AT4040. Read on 08.02.2022. https://www.thomann.de/fi/audiotechnica_at4040.htm

Thomann. N.d. Electro-Voice RE20. Read on 08.02.2022. https://www.thomann.de/fi/ev_re20_microphone.htm

Thomann. N.d. Neumann TLM 102. Read on 25.02.2022. https://www.thomann.de/fi/neumann_tlm_102_nickel.htm

Thomann. N.d. Neumann TLM 103. Read on 08.02.2022. https://www.thomann.de/fi/neumann_tlm_103.htm

Thomann. N.d. Rode NT1-A Complete Vocal Bundle. Read on 08.02.2022. https://www.thomann.de/fi/rode_nt1a_complete_vocal_bundle.htm

Thomann. N.d. Sennheiser MKH 416 P48. Read on 08.02.2022. https://www.thomann.de/fi/sennheiser_mkh416p48u3.htm

Thomann. N.d. Shure SM 7 B. Read on 08.02.2022. https://www.thomann.de/fi/shure_sm_7b_studiomikro.htm

Thornton, M. N.d. Sound on Sound. Using Fades & Crossfades. Read on 12.04.2022. https://www.soundonsound.com/techniques/using-fades-cross-fades?amp

TritonAudio. N.d. TritonAudio FetHead. Read on 08.02.2022. https://www.tritonaudio.com/fethead

Winer, E. 2016. Acoustic Treatment and Design for Recording Studios and Listening Rooms. Updated on 09.05.2016. Read on 13.02.2022. http://ethanwiner.com/acoustics.html

Vinnie. Home Studio Expert. 2021. How Quiet Should a Recording Studio Be? Updated on 16.09.2021. Read on 13.02.2022. https://homestudioexpert.com/how-quiet-should-a-recording-studio-be/

Warehouse-Lighting.com. N.d. Importance of Proper Lighting While on the Computer. Read on 12.04.2022. https://www.warehouse-lighting.com/blogs/lighting-resources-education/importance-of-proper-lighting-while-on-the-computer

WebMD. 2020. What Is Misophonia? Medically reviewed by Arefa Cassoobhoy, MD, MPH on 13.12.2020. Read on 28.02.2022. https://www.webmd.com/mental-health/what-is-misophonia

Winstead, S. 2021. My eLearning World. How to Get Rid of Mouth Sounds in Your Vocal Recordings — 10 Tips by Scott Winstead. Read on 12.04.2022. https://myelearningworld.com/avoid-mouth-noises-when-recording/

Wreglesworth, R. N.d. Musicians HQ. XLR or USB Microphone for Vocals? – The 6 Reasons to Choose XLR. Read on 12.04.2022. https://musicianshq.com/the-6-reasons-to-get-an-xlr-microphone-over-usb/

Özdemir, B. N.d. Alarm Journal. What Time Do People Go To Bed? (Data-Driven Insights). Read on 14.04.2022. https://alarmjournal.com/what-time-do-people-go-to-bed/