Non-diegetic sound effects

Keywords: Diegesis, acousmêtre, synchresis, sound effect, film score

Abstract
Most films tell a story. This story takes place in a constructed world that only exists in each specific film. It makes no difference if the story is set in a fantasy world, a historical or a future world, or even if it’s a documentary that depicts reality. Every film still creates it’s own filmic world, its diegesis. Everything which happens inside this world is called diegetic and what happens (in the movie) outside this world is called non-diegetic. Traditional film music and voice-over narration are typical examples of non-diegetic sounds. The characters in the film are unaware of them, because they don’t exist in the same world. These sounds are messages from the filmmaker directly to his/her audience. Music played inside the film’s world, for example by visible musicians or from a radio seen on screen, is diegetic, as is dialogue and sound effects. The characters in the film are meant to be aware of these sounds. Whether the actors heard these sounds while shooting the scenes or if they where added during sound editing, they influence the audience’s interpretation of the characters, the situation and the narrative.

Diegetic or non-diegetic, a non-issue?
In his article Acoustics of the soul (2007), Sound editor and re-recording mixer Randy Thom writes:

I've never heard anyone ask a director a question like, "is this gunshot diegetic?"
or "is this saxophone solo diegetic?" I suspect one reason is that there are more straightforward ways to ask such questions when working on a film: "Does Jim hear the gunshot?" "Does Angela hear the saxophone?" !

In my own experience, filmmakers don’t use the terms diegetic or non-diegetic, but we’re certainly aware of the issue and it’s implications. Some film theorists on the other hand have begun dismissing the term because they claim that everything in the film is part of the diegesis. The argument for this is that the film is really constructed by the viewer, not by the filmmaker(s) (Stockfeld, 1996). So whether the term itself is obsolete or not, I would like to argue that it’s implications for the narrative remain as relevant as ever. If
the audience interprets a sound as diegetic, they will understand the characters’ actions in relation to the soundscape that surrounds them. If they interpret it as non-diegetic, they will consider it as information from the filmmaker directly to them. It’s a hint from the director without the characters being aware of it. Moving a sound between these two audio modes can be a very powerful narrative tool.

The acousmêtre and the diegesis

In his book *Audio Vision, Sound on Screen*, Michel Chion (1999) establishes the term ‘audio-visual contract’. This contract can be seen as an agreement between filmmaker and audience about what rules apply to the use of sound and image in the film. The implication of his argument is that the audience accepts that what’s happening on screen is ‘reality’ for the duration of the film and that everything outside the film, is irrelevant. The contract is based on the fact that the audience knows how a movie narrative works and what rules filmmakers usually follow. They are also aware of different genres and the different demands and opportunities that come with them. Of course, such knowledge is learned and constantly evolving.

Filmic innovations such as close-ups, camera movements, parallel action, sound film and colour film had to be learned by the audience before the filmmakers could develop them as integral parts of their stories. Now that we face 3D, three-dimensional images, further conventions may need to be created. It will take some time before everyone agrees on how this technology can be best used and what its implications are for the narrative.

If it proves difficult to find a natural integrated function for the depth dimension of the image, 3D runs the risk of ending up on a shelf next to Smell-o-Vision, Odorama and Sensurround in film museums.

The audio visual contract also includes an agreement that the sounds we hear in the film belong to it, and even more specifically, that the sounds are actually caused by the events we see on screen. In Audio Vision (Chion 1994) Writes:

> *Syncretism (a word I have forged by combining synchronism and synthesis)* is the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time.

The spectators themselves make the connection between the visual event and the sound that is played with it. This means that the average spectator of a reasonably well-made film doesn’t question the sounds she/he hears. If the sounds seem strange, they say something about the object or event that caused them, but they don’t say anything about...
the sounds as such. The sounds are no longer independent events, but something that is tied to the image. This is what makes creative sound editing possible. You have to move very far from realism when choosing a sound for a certain visual event, to break the synchresis.

In Jacques Tati’s (1967) film *Play Time* there is a scene where M Hulot (played by Tati himself) visits a furniture exhibition. Among the stands there’s one that shows ‘the silent door.’ Although it’s slammed shut several times, it never makes a sound. It is probably unlikely that a door can work that way in the real world, but you don’t question it in this case. Silence is also an audio feature and in Tati/Hulot’s world; it is plausible.

Another part of the agreement is that the score and the voice-over narration belong to the film, but exist outside the diegesis. We don’t expect to see the orchestra or the narrator, nor do we expect any of the characters to react to or comment on them.

If the source of music (by showing musicians, a radio, or a location where music is usually heard) is not shown, the audience will perceive it as being non-diegetic. In the case of the narration, it is sufficient to follow the convention of how a voice-over should sound, and perhaps also how it’s normally used, for it to be perceived as something that exists outside the diegesis.

Chion also launches the term *acousmêtre*, the sound being or sound object, as the name of an object outside of the picture frame, which is the source of a sound we hear. It’s a disembodied sound that’s waiting to be embodied through synchresis. He argues that the acousmêtre creates a tension based on an uncertainty in the audience as to whether they are going to see the sound source, or not. One cannot avoid reflecting on what it might look like, whether it is a human being whose voice we hear, or if it is a known or unknown object. According to Chion this also applies to narration. However, he doesn’t mention the music score in this context. This implies that the acousmêtre does not apply to non-diegetic sounds. I would like to argue that there is a distinction in the audience’s perception between the sounds that are included in the diegesis, and those that are not. If I’m right, neither the narration nor the score are acousmetric, that is, the audience does not expect to see either the film music orchestra or the narrator appear on screen.

Chion’s classic examples of acousmêtre are Dr. Mabuse’s voice in the film *The Testament of Dr. Mabuse* (Lang, 1933) and the Wizard’s voice in The *Wizard of Oz* (Fleming, 1939). In both cases the tension comes from the fact that we initially don’t see the origin of the voices, and thus they are perceived initially as akin to narration. We do however see the curtain behind which the speakers are said to be; the origin of the voice
has a physical space in the diegesis. This is not the case with a voice-over.

**Sound that comments on or participates in the story?**

According to the usual definition, sound effects, short sounds that illustrate events that occur on screen, are definitely diegetic, because one actually sees the sound source. This fact doesn’t change even if the sounds are completely artificial, for example sounds of punches in fight scenes. The actors don’t actually hit each other with any force, and the sound effect can be made from a combination of sounds from different sources. Another example is sounds for computer-generated images. They don’t exist in the physical world and thus can’t make any sounds at all. Most members of an audience are aware of this, but nevertheless, through synchresis, perceive these sounds as coming from the events on screen. The light sabres of *Star Wars* (Lucas, 1977) are a good example of this. However unrealistic, their sound still changes the way we see the sabres, the characters using them, and their actions.

Naturalistic background sounds and atmospheres such as traffic noise, bird song and wind sounds are also diegetic, especially if they’re realistic in relation to the scene’s location. Regardless of what’s on screen, it’s likely that these sounds come from somewhere in the background or off camera. Even if the background is hidden in darkness or completely out of focus, we know that the scene is set in a city or in a meadow, and thus the background sounds are diegetic and as a result, the characters are meant to be aware of them. If the scene is set in outer space, these same bird sounds and traffic noise will have a very different function. Now they’re not realistic in relation to the setting of the scene, and thus non-diegetic.

If diegetic and non-diegetic represent opposite poles on a vertical axis that describes a sound's authority, there’s a second, horizontal axis, with the categories Wingstedt (2008) establishes to describe the narrative functions of film music. He divides them into:

- emotive function (to convey emotions);
- informative function (including information about era, cultural space/environment, and social status);
- descriptive function (describing movement and size);
- guiding function (leading the eye or the thought);
- temporal function (creating continuity, define structure and form - in a scene, or in
a film as a whole);
- rhetorical function (contrast the story with well-known pieces of music.

The same categorization can also be used for sounds that aren’t music, and even for dialogue.
The fact that dialogue usually is used as a conveyor of information doesn’t mean that it can’t have one or more of the other functions as well. The same applies to background sounds and sound effects. These can sometimes approach music when they have an emotive function. Aggressive and unpleasant sounds, for example, often contain rhythmic components, low frequency (tonal) or piercing metallic sound (dissonance).
Like score, a sound can change function and serve different narrative purposes during a scene or throughout a film.

In *Atlantis* (Besson, 1991) the sound editor uses, among other sounds, murmur from a concert audience and bird song to create the atmosphere in underwater scenes. These sounds serve as a commentary to the pictures, suggesting to the viewers a similarity between the underwater world and the one above the surface. The sounds thus state, in a humorous way, that humans and other land creatures in some ways are similar to the ones living in the sea. This is an obvious rhetorical function.

These sound effects and background sounds are edited so that they follow the movements of the fish, but they obviously don’t originate from the image. These everyday sounds are perceived as something outside the diegesis. Although they are completely realistic, they are non-diegetic. The synchresis is broken. No one expects to see birds or a theatre audience appear in this underwater world. Instead we interpret the sounds as messages directly from the filmmaker, outside the diegesis.

This example shows how far you have to go to make a realistic sound non-diegetic in a realistic environment. If the sound and the image have a closer link, our consciousness tries to connect them. The humour in this case is a result of these futile attempts by our brains. It also shows that non-diegetic sound effects are possible and can be used as a creative narrative tool.

Another example of non-diegetic sound effects is the synthetic helicopter sound in the beginning of *Apocalypse Now* (1979) The sound is obviously synthetic, but at the same time it’s clearly connected to the helicopters. It’s a sound that *means* helicopter but clearly comes from another source.

So, in this context how should we categorize such subjective sound? A sound that we
claim to be subjective is by definition diegetic, even if it’s only from one character’s perspective. It is non-diegetic in relation to all the others, who don’t hear this particular sound that perhaps only exists in the imagination of one of the characters. To establish a point of audition the filmmaker first of all has to make the audience understand which one of the character’s perceptions we are sharing. When this is accomplished, the sound is definitely diegetic. One could establish the term ‘subjective diegesis’ for such cases. In the film \textit{Vildängel} (Engberg, 1997) I illustrated a climb up a high ladder on a crane by using metallic sounds that were exaggerated to describe the vertigo of one of the characters. The sounds themselves are sort of realistic, but played rather loud in the mix. To underline the fact that it’s this particular character who’s experiencing this fear and discomfort, these sounds are only heard when he moves. When his protagonist moves, everything sounds more or less normal. This is an example of how subjective diegesis can work in a narrative. Another example is the concert sequence in \textit{Shine} (Hicks, 1996). The music changes from sounding like an ordinary concert to something radically different when it’s filtered through the mind of the main character. What the audience at the concert is hearing is of course diegetic, but what the main character is hearing is subjective diegetic. He reacts to what he hears, and we understand his acting in relation to his distorted perception of the music.

Moving the sounds between diegetic (subjective or objective) and non-diegetic is thus to change them from participating in the action to commenting on it.

\textbf{The voice-over}

On the non-diegetic level where the music score resides, there’s only one more player, the voice-over narration. We don’t expect the characters of the films to comment on or react to any of them.

As Kozloff (1988) writes, many film scholars and filmmakers traditionally have dismissed voice-over narration for being non-cinematic. She argues that this is, to a certain degree, is the consequence of their general aversion towards dialogue in film. She quotes, among others, William Goldman (1983) who writes ‘In a movie you don’t tell people things, you show people things’.

This leads to a kind of hierarchy between what we see in the picture and what is said on the non-diegetic level. The same is probably also true between what we \textit{hear} from the screen and the non-diegetic messages. Music score and narration are perceived as more manipulative than what’s seen and heard in the diegesis. We see with our own eyes and
are able to form our own views based on that, the score and the narrator make categorical statements that we don’t necessarily share. This seems to apply no matter how manipulated and cropped the image may be and no matter how artificial the sounds associated with them are.

The godlike property of voice-over is a very powerful narrative tool if used wisely. The British documentary filmmaker Sir David Attenborough has found his own way of exploiting this. He may start a sentence as the narrator, continue it in sync, on screen in a jungle in the Amazon, carry on over a close-up of an animal there, and then finish the sentence in sync sitting on an iceberg in the Arctic. There’s no change in his voice, despite the radical changes in the environments. His voice remains that of a narrator. Whether he is on screen or not, and wherever he is in space and time, he is indisputable. It is a very efficient way to display and maintain authority.

To obtain this authority there are strong conventions on how a voice-over should be recorded. It must be read in a neutral tone of voice, the microphone has to be close to the speaker’s mouth to create a sense of intimacy, and it must be recorded without room acoustics or ambience. That is, there can’t be any other information other than the actual meaning of the words that might diffuse the message or cause doubt in the viewer. The voiceover is like a voice inside the listener's own head, and we usually don’t question such a voice.

**Music score**

Non-diegetic music is a message from the filmmaker directly to the audience. It serves as a handrail for the audience, so that they know where the story is going. One of the points of score is that the characters in the film are not aware of it. But, how can the audience know that it is music that they’re hearing, if the music doesn’t sound like ‘music’?

There are many examples of how composers have tried to find the boundaries of music, by using ‘unmusical’, natural or fabricated sounds in their compositions. Electroacoustic music and noise music are examples of this.

If a sound is transformed into, or functions as music (in the spectator's view), the expectation of seeing the sound source disappears. At the same time the sound's ability to influence our perception of the event or location changes. That is, the sound is also transformed from diegetic to non-diegetic.

But what happens if I want to use electro-acoustic music diegetically in a scene? Do I have to treat it differently than, let’s say, a piece of orchestral music, in order to be sure it
will be perceived that way? There must be quite a risk that a large part of the audience perceives that kind of music as sound effects. As this appears to be significant, it may be a problem in this particular case.

What happens if you move between these two modes: between diegetic sound that creates certain expectations in the spectator, and non-diegetic that create others? Will this make the audience aware of the film form by breaking the rules and changing their perception of the movie? Is it similar to when an actor looks into the camera lens, or when the microphone dips into the frame?

A sound editor or a dubbing mixer may use an uncertainty about this to manipulate the authority of different sounds, and thereby the audiences’ perception of a scene or a location.

There’s a scene in *Seven* (Fincher, 1995) in which two police officers enter a killer's apartment. The audience is presented with a complex array of different sounds. Which ones of all the sounds heard are music, and which ones are diegetic sound effects or atmospheres? Are some of them even subjective sound and if so, from which one of the characters’ point of audition? Maybe some of the sounds move between these different states and thus sharpen our attention and raise the tension? The images are also playing along in this game. The camera moves and the perspective shifts from objective to subjective. Sometimes the image is the point of view from one of the police officers' perspective, sometimes it’s objective, and sometimes it could be the point of view from a third person who is hiding in the dark. This type of dynamic cinematography helps to increase the feeling of uncertainty and makes the viewers uncomfortable, just like the characters on screen, and that’s exactly what this scene is about. The constant shift of the diegetic level of the different sounds also forces the spectators constantly to re-evaluate what they see and what they hear, and the relationship between these sensory inputs. I doubt however that these bold choices made by the sound crew and the composer would have been possible without this kind of visual language.

**Sound effects turn into music that becomes FX**

It is relatively easy to change music so that it takes on the narrative properties of sound effects. The most obvious way is to make such music diegetic. In my own experience it diminishes the emotive function as well as the descriptive, while it has less impact on the informational and rhetorical functions. The temporal function doesn’t seem to be affected at all.
Imagine playing the *Jaws* theme through a small radio inside the fishing boat when the shark attacks, instead of playing it non-diegetically. It would alter the narrative effect quite dramatically.

It’s hard to tell why viewers interpret diegetic music differently how they interpret non-diegetic. It can’t only be the technical quality that matters. The fact that we often cut out the low frequencies and exaggerate the mid to create radio or loudspeaker sound, can’t be decisive. Neither can the comparatively lower level at which diegetic music often is mixed. Could the important factor be that the music resonates in the same acoustic space as the dialogue? It loses its godlike quality and starts to participate in the narrative rather than commenting on it.

The American sound designer and film editor Walter Murch made a famous experiment when he created the musical backgrounds in *American Graffiti* (Lucas, 1973). By playing music through loudspeakers in a car park, and then recording that sound from a distance, the music was heard with a natural and varied outdoor timbre. He named the procedure *worldizing*, (Jarrett (2/4 2012). This meant that the music now could be perceived as existing in the diegesis, among the characters. In this film, the music constantly changes between diegetic and non-diegetic. It also changes its narrative function between all the previously mentioned categories. It’s interesting to see that these changes take place and work without the need for an explanation. Of course, the informative function is only important at the beginning of the film, at least in terms of age and social status. Once we’ve understood that message, we don’t need to be reminded again. A less obvious statement is that the entire town where the story is set listens to the same radio station. This message grows and becomes stronger as the film continues. The music also serves other purposes. The song lyrics often comment on events in the plot, ironically or not (rhetorical function), they create continuity in scene-shifts (temporal function) and some songs are tied to some of the characters (descriptive function).

I used this method when I was sound editing a TV-series. The most important concept for the series was that none of the settings in the story should be described as being pleasant. Every office, machine hall and dormitory had to be intimidating, dangerous or hostile. On the other hand the director had chosen beautiful music to be used diegetically in most of the scenes. So, I was left with the task to remove or distort all the emotive properties from Wilhelm Peterson-Berger and Hugo Alfvén’s music. I ended up playing it through a small loudspeaker at a very high level, into a megaphone at a volume close to feedback. It was then rerecorded at the other end of a corridor outside my studio. The
result sounded horrible. But, since the emotional content of the music is so strong, it
needed this much distortion to become appropriately diminished. It became quite clear to
me that the narrative function that seemed to survive regardless was the temporal. I could
still use the music to create or break continuity in and between scenes, even with the most
trashed versions of it.

Another way to change the typical musical features of music is to remove the rhythm
and/or the melody. A long chord or single notes are examples of this. It’s possible to
slowly fade in and fade out a sound like that in a mix, without anyone noticing its musical
origin. As long as you leave out the attack from the instrument(s) it’s quite
inconspicuous.

These kinds of sounds also exist outside the world of music. Wind, traffic ambience and
gine noise can have tonal qualities. The Swedish sound artist Mikael Strömberg has
even tried to determine the tonal key of the general ambience of some cities and towns. In
Ljudbiblioteket (2008) he writes about the sounds of S:t Jean de Luz, a french coastal
town. ‘A fragile shimmering soundscape tuned in fourths and fifths’ (my translation).

In film sound editing it is common practice to combine ambience sounds with low
frequency sounds, for example from a synthesizer.

If you start adding sounds with tonal qualities and there’s also score in the scene you
have to decide if the added sounds should be harmonious or dissonant in relation to the
music. But whether or not you add tonal sounds to your ambiances, the perceived pitch of
the sound will be added to the music and influence it, just as the music will affect the
other sounds. The same applies to rhythmic sound effects. The perception of rhythm in
machines and footsteps changes the music, and is changed by it. The boundary between
the diegetic and the non-diegetic thus becomes very diffuse.

Let me illustrate this with an example. If the audience perceive the backgrounds in the
corridors of The Night Watch (Bornedal, 1994) as being atmosphere sounds, then they are
diegetic, but if they identify it as music, then they are non-diegetic, regardless of how the
sounds were made. Each listener decides for him or herself.

If the sounds are perceived as diegetic realistic atmospheres, they become an acousmêtre.
The sounds create a tension and an anticipation as to whether the sound object is on
screen or will appear on it or not. Maybe it’s noise from the ventilation duct in the ceiling
I hear, or there may be a potential danger around the next corner that I haven’t yet seen.
If however, the audience perceives the sound as non-diegetic music, it creates different
expectations. Now it’s the filmmaker who’s telling us that there are dangers that the
characters are not aware of. Or, maybe the music indicates how the character perceives the situation. In that case, we must consider whether that view is plausible or not. S/he could very well be afraid of the dark in a completely harmless place.

Audiences have acquired knowledge and tools to decode all possible kinds of messages in films. They probably have no problems understanding different kinds of music, or various kinds of other sounds.

So, what happens in the narrative when music goes from being diegetic to be non-diegetic?

Non-diegetic music exists on the highest authority level in the film. It is a straight path from the filmmaker to the audience's hearts and minds.

When diegetic music becomes non-diegetic, it is a way for the filmmaker to tell the audience that the message or emotional state conveyed in this piece of music chosen by this specific character has more relevance than they know. This is very elegantly demonstrated in the French film *Diva* (Beineix, 1981). The secret recording made by the main character at a Paris concert becomes part of the non-diegetic score. It’s a way to elevate this particular piece of music and its connotations to absolute truth. The same thing happens in *Apocalypse Now* when Wagner’s music played from the American helicopter, turns non-diegetic. It becomes a strong moral statement about the Vietnam war.

But what happens if we do the opposite? If the non-diegetic music suddenly starts playing diegetically, what does this mean? Maybe it’s an invitation from the filmmaker to the characters to actually comment on the indisputable. They may dismiss the score or they may accept it, such as the sheriff in *Blazing Saddles* (1974) who rides across the prairie and suddenly passes Count Basie and his orchestra playing the score. It’s a good joke and a bold move by the filmmaker to allow the characters of his own making to scrutinize his choice of music.

By whatever expression one chooses to call it, I am convinced that the question of diegetic or non-diegetic is as relevant as ever. It has implications for all auditory aspects of film narrative and it can be used in conjunction with several other analytical tools for film sound.

Klas Dykhoff is a sound editor and Professor of Film Sound at the Stockholm Academy of Dramatic Arts. Klas.dykhoff@gmail.com
Sources


Besson, Luc (1991), Atlantis, France, Gaumont, Cecchi Gori Group Tiger

Bornedal, Ole (1994) Nattevagten, Denmark, Thura Film.


Coppola, Francis Ford, (1979) Apocalypse Now, USA, Zoetrope Studios

Ebert, Roger, Why 3D doesn't work and never will. Case closed.

Engberg, Christer (1997), Vildängel, Sweden, Filmpool Nord, Giraff Film AB, Kulturskolan Rosteriet


Goldman, William (1983), Adventures in the Screen Trade, New York, Warner

Hicks, Scott, (1996), Shine, Australia, Jane Scott

Jarrett, Michael Interview Sound Doctrine: An Interview with Walter Murch
http://www2.yk.psu.edu/~jwj3/murchfq.htm (2/4 2012)


Lang, Fritz (1933), Das Testament des Dr. Mabuse, Germany, Nero Film AG.


Lucas, George (1976) Star Wars, USA, Lucasfilm, Twentieth Century Fox Film Corporation
Spielberg, Steven, (1975) Jaws, USA, Zanuck/Brown Productions, Universal Pictures


Tati, Jaques (1967), *Play Time*, France, Jolly Film, Spectra Film


Wikström, Pontus (2002) *Svenska Slut*, Giraff Film AB, Filmpool Nord Sveriges Television


---

1 OFFSCREEN :: Vol. 11, Nos. 8-9, Aug/Sept 2007
2 Films that include smell for added sensation
3 A sub sonic sound system for shaking the audience
4 Walter Murch shares some interesting thoughts about 3D in Roger Eberts’ article http://blogs.suntimes.com/ebert/2011/01/post_4.html
5 Wikström, Pontus (2002) *Svenska Slut*, Giraff Film AB, Filmpool Nord Sveriges Television