6EWgjJtoiW6R

# Navigating the Radio Archive: Segmentation and analysis of Swedish broadcasting data

### Johan Malmstedt

Umeå University

[ ©. Published by De Gruyter in cooperation with the University of Luxembourg Centre for Contemporary and Digital History. This is an Open Access article distributed under the terms of the [Creative Commons Attribution License CC-BY-NC-ND](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Radio Studies, Public Service Broadcasting, Audio Analysis, Media History

This work aims to explore the relationship between sound archives and historiography, focusing on the Swedish case of mass media archiving from the 1980s. The study investigates the sonic content of public service radio changes during the introduction of commercial broadcasting, using computational methods. In the analysis, two of the most popular public service broadcasting channels, P1 and P3 are compared. By analyzing 1,600 hours of radio data, the paper reveals a shift in the number of detected occurrences, with varying sonic sequences that reflect the overall structure of the broadcast. Although the sample sizes are small, the findings show a correlation between object detection and dimension reduction, suggesting and increasing attention to content variation.

The paper contributes to the understanding of historical radio data, offering preprocessing and segmentation methods for working with cultural audio data. It also emphasizes the methodological implications of combining dimensionality reduction and object classification approaches, demonstrating the value of using pretrained and untrained algorithms together for a comprehensive understanding of the local and fine-grained aspects of audio data. The necessity of such an approach springs from the oceanic extent of content in the radio archive. Thus, the article suggests new ways to navigate the radio archive

## Introduction

Sound archives are peculiar features in the 21st century – extensive and vast, yet mute and inaccessible accumulations of seemingly arbitrary signals from the past. But what is the relationship between signals and history?

This work is an experimental attempt to plug the signal output back into historiography. Technical, ethical, and epistemological factors have long since short-circuited the historian’s relationship to acoustic signals. As Friedrich Kittler notoriously stated already in 1999, “discourse analysis cannot be applied to sound archives or towers of film rolls” (*Kittler 1999*). Yet, due to a renewed interest in broadcasting material, audio data is achieving an increasingly significant status within the historical discipline (*Bösch, Bourdon, Meyen, Spigel, von Hodenberg 2012*). Sound archives of multifarious natures are expanding across the world and digitization is running ever faster. Their oceanic density of content requires new methods of navigation.

Let us consider the Swedish case; during the 1970s, discussions within the archive and library sector were centered around ‘new’ media forms within the public service realm. Television and radio are argued to “be retained to an extent corresponding somewhat to that applicable in the case of printed material” (14492245/95XIHHYV). This idea was not unique at the time; broadcasting media was already considered an informational force to be reckoned with. The Swedish case is particularly interesting because it is one of the first attempts to act on this insight. In 1979, the doors opened to a new type of archive – a mass media archive. The results invite a double entendre of the word ‘mass’; not only is the content supposedly intended for the so-called masses, but the archive is also massive. Storing audio to the same extent as text means establishing a legal deposit law that enforces the storage of not just ‘certain’ content, but all content. In terms of audio media, this entails, among other things, several radio channels, broadcasting around the clock, seven days a week. Unsurprisingly, the archive quickly grew to be one of the largest of its kind. At this particular moment of writing, March 11th, it hosts 634746 radio recordings alone, most spanning several hours each.

This poses a challenge for the historian. Philosophy of history has repeatedly stressed the text-oriented nature of the subject. Hayden White famously made this point central to his project (*White 1988*). However, theory is not needed to reveal this historiographical fact. The very word itself makes clear that it is a profession to the ‘writing’ of narratives. The historian writes about what has already been written – or at least this was the case for most of historiography. Today this case is rapidly changing. Not only are historians, as mentioned above, becoming increasingly interested in non-textual sources. At a more fundamental level, the culture of the past is undergoing a tectonic shift in modality. As White published on the topic, back in 1980s, textual materials, papers, and manuscripts still constituted a triumphantly towering source material from our past. Today audiovisual signal data grows feverishly to challenging expanses. Before long, this type of material will be the principal format in which the past reaches us.

The signal is imposing its nature on cultural studies and the human sciences. This means that the humanist must engage with what Umberto Eco once described as the “lower threshold of semiotics” (*Eco 1976*). In particular, for this case, historians will have to become familiar with audio data. To advance in this direction, my article combines a historical analysis of Swedish radio with introductory remarks on audio analysis. The aim is to provide insight into the stylistic development of public service broadcasting (PSB) while also offering an introduction to current audio processing methods. The first segment explores a contemporary radio sample, demonstrating the effective features for this specific type of source material. Unlike other fields with acoustic source data, such as musicology and bioacoustics, the historical sciences still lack both methods and understanding of the sonic data. Therefore, the focus of the first part is to provide some suggestions for the navigation of historical radio data. The hermeneutics layer provides the reader insight into the concrete measures applied, while the narrative layer demonstrates the main points of the excursion.

The second part of the analysis deploys the method advanced in the first section on a dataset of 1600 hours of radio data to sort and analyze the content. This segment focuses on a comparative analysis of pretrained and non-pretrained methods, exploring questions related to variation and stylistic changes in PSB towards the end of the past century. The general research question of the article is thus:

How did the sonic style of Swedish public service radio change before and after the introduction of commercial broadcasting, and how can this be explored computationally?

Though valuable both from the perspective of radio history in particular and from the larger scope of media ecology, the stylistic impact of commercial radio has received too little attention. Studying how the actual content of PSB changes during times of shifting competitive conditions may not only reveal insights into the development of Swedish radio but grant one further step in understanding the relationship between the circumstances of media production and the actual sonic output. To better understand the significance of this endeavor, the case of Sveriges Radio (SR) is principal. Until the early 1990s, the concept of radio style was limited to matters of separate channels within the monopoly. For example, P1 was a speech-oriented, informational channel, while P3 was a popular music channel, and P2 was oriented around classical music and jazz. The tempo, distribution of music and speech, tonality, and overall aesthetics of the channels were, at least during the first part of the second half of the century, clearly distinguished. The stylistic matter was chiefly a question of internal balance between these channels. However, during the 1990s, this position was subject to change. In 1993, SR instigated an internal “product development group”, tasked to determine the PSB “identity”. The cause was the “new competitive landscape”, a euphemism for the commercialization of the airwaves that took place the same year. From September 1993, broadcasting would be effectively available to any commercial broadcaster with a license.

The impact of such structural reorganization in broadcasting remains a contentious topic. On one hand, there is a belief that these changes should not disrupt the economically independent institution of PSB. On the other hand, the funding agreement for (SR) emphasizes the importance of maintaining a public audience. A significant loss of listeners to commercial alternatives could jeopardize the stability of PSB. Consequently, the issue of audience appeal is increasingly viewed through the prism of commercial competition. This shift necessitates a reevaluation of PSB’s style, as it’s not just the existence of SR at stake, but also key considerations in media studies. The critical analysis of commercial radio, a recurring theme in critical radio studies, is well-articulated by scholars like Wolfgang Hagen and David Hendy (!Hagen, 2013, Hendy, 2003). While my work does not dispute their critiques of the monotonous and unoriginal output of American commercial broadcasting, it’s worth noting that these arguments often presuppose the existence of ‘non-commercial radio’. As these discussions extend beyond economic and organizational structures to include the ‘programming of self-identity’, it becomes crucial to examine a case like Sweden’s to identify distinctive traits of non-commercial broadcasting.

## Audio data and historiography

Audio data is not inherently narrative or discursive, though they often come to be perceived as such through various transformative processes, including sonification, visualization, and the use of machine learning algorithms for speech-to-text conversion. Fundamentally, audio data consist of sonically oriented, time-discrete values that represent acoustic information. This notion has been the subject of much scholarly debate, highlighting its paradoxical nature. Prominent figures like radio scholar Wolfgang Hagen and media archaeologist Wolfgang Ernst have delved into the epistemological complexities surrounding audio data. Ernst, in particular, has elucidated how audio files are fundamentally non-sonic in their dormant state, being mere numerical entities that become interpretable to humans only through signal processing (*Ernst, Parikka 2013*). We are, therefore, engaging with ephemeral representations that can be perceived in unconventional ways, such as lists, three-dimensional shapes, and points in Euclidean space. To delve deeper into this concept, let’s consider the application of the audio processing library Librosa and examine a test file:

This is an audio file represented in the so-called time-domain. It is commonly used for skeuomorphic purposes and might therefore be recognizable for the broader audience. Nevertheless, its smooth, suggestive surface hides a more complex truth. By zooming in on the visualization, we can come close to the bare bones of pulse code modulation, which is a means of measuring the soundwave at very small intervals. The result corresponds to the fluctuation of air engendered by this specific sound. While PCM remains a significant method for storing digital audio, its raw form is largely indecipherable to human perception. Visually, one might estimate fluctuations in dynamics throughout the audio file, but compared to the auditory experience, such a representation is comparatively sparse in information. To extract meaningful insights regarding pitch and timbre, we must resort to Fourier transformations. Friedrich Kittler, in his essay on real-time analysis, aptly describes this method as a “mathematical magic trick from the 1820s that has since become indispensable for the computer age”, facilitating the transition from whole to real numbers, from combinatorics to calculus (*Kittler, Winthrop-Young 2017*). This process unveils the mathematical harmonies that constitute a sound’s unique character. The illustration below offers a straightforward demonstration of how Fourier’s technique transforms data from the time domain into the frequency domain.

The spectrum displays the distribution of frequencies, or what we roughly can consider as pitches. Visually, this tells us significantly more about the character of the audio than the waveform above. Regrettably, this comes at the cost of temporality – there is no time axis. Instead, this method displays a general distribution of frequencies for a specific amount of time. This plot displays the distribution in the first “frame” of the file - a significant concept in audio processing. In the following analysis, we shall employ the audio processing standard, which is a frame length of 2048, and a sampling rate of 44.1 kHz, which means that the actual sample window is about 46.4 milliseconds. The image above thus tells us something about the tonality of the first half second of the audio file.

To further determine the content of an audio file, we can proceed in several ways. One could simply describe it as a 3-minute excerpt of American radio from 1936. It is the dramatic first minutes of the classic detective story “The Thin Man”, starring Myrna Loy and William Powell. Though these names might still have cultural resonance in some places, the description actually reveals little about the sonic content. Alternatively, one could begin by processing and analyzing the signal information, which would require a more advanced application of signal processing. Fortunately, audio signal processing has been developed in a variety of acoustically-oriented fields. For example, phonetics employ processing methods for precise, small-scale analysis of speech sounds, where musicologists have adopted a sophisticated use of digital methods, translating classical music analysis into computational matters. However, none of these approaches fully address the challenge of audio data spanning hundreds, or thousands, of hours. This challange is more common to computational bioacoustics, where specific animal vocalizations are to be detected in vast recordings. As a result, automated content segmentation and selection have become crucial under such circumstances, resulting in a variety of methods for detecting animal activity in large datasets. This comes methodologically closer to radio studies but differs in the sense that the aim usually concerns the targeting of rare and specific events in relatively information-sparse content (*Bravo Sanchez, Hossain, English, Moore 2021*).

Recent years have seen a growing number of endeavors to integrate computational audio analysis methods into humanities research. A notable pioneer in this field is The Institute on High-Performance Sound Technologies for Access and Scholarship (HiPSTAS), established in 2013. HiPSTAS has been instrumental in breaking new ground in humanities scholarship by applying computational audio analysis techniques (*Clement 2015*). Scholars such as Iben Have and Kenneth Enevoldsen from Aarhus University, and Golo Föllmer from Halle University, have made significant strides in applying these methods specifically to radio studies (*Have, Enevoldsen 2021*, 14492245/D4BQ5Q8V). My work seeks to build upon and extend these foundational efforts. Specifically, I intend to leverage the unique capabilities of this journal to engage the reader not just visually, but also aurally, by inviting them to participate in the experiments as listeners. The methodology employed in the following section will be intentionally simple. This approach serves a dual purpose: firstly, to ensure methodological transparency and simplicity, which I hope will encourage further exploration and refinement by others, and secondly, to facilitate a deeper discussion about the essence of historical sounds.

My argument is that the historian working with mass media data, and radio recordings in particular, is dealing with a specific type of source material. This is not to reduce the contextual complexity of audiovisual material. Rather, I simply want to address the status of signal-based cultural heritage archives and their inherent challenges. Even if many of these challenges overlap with bioacoustics, there are distinct epistemological issues of radio data. Within broadcasting, silent segments, which can aid in data segmentation, are scarce. Radio is characterized by its continuities, with countless hours of uninterrupted content, necessitating that segmentation be arbitrarily determined by the researcher. This constitutes a departure from the computational study of music as well. Unlike music, which typically has clear beginnings and endings, radio broadcasts often lack these demarcations. Therefore, the initial part of this article will focus on exploring potential solutions to the challenge of segmenting continuous radio data.

To explore the function of audio segmentation, we can initially test the predeveloped method provided by librosa. The onset.detect function will look for significant changes in amplitude and make a time stamp wherever they appear:

The method certainly detects events in the audio file, but they appear with a very high regularity. This might be appropriate for certain types of analysis but considering that we are dealing with more than a thousand hours of radio, this density will not be very beneficial. Instead, a higher level of discrimination is needed. Librosa offers a solution to this problem. The reader can experiment with this by going to the hermeneutics layer above and adding the argument “delta” to the function “librosa.onset.onset\_detect”. Delta determines the threshold level for at what point value change shall be considered as significant.

However, as delta approaches 0.3, discrimination begins to reach the limits of effectiveness. Only a few detections remain, but not at time stamps that correspond helpfully to the actual content. Several detections are still made in the same speech sequence, whilst musical elements are lost. This does not entail that the method is invalid. Considered as an arousal of activity in the audible spectrum, delimited by the restabilizing of ‘silence’, each of these chopped-up sentences are sonic events. However, this is not the only definition of a sonic event. In fact, my proposal here is that this is a rather inappropriate definition of a sonic event in the context of radio. Instead, I will attempt to propose a simple alternative.

In the endeavor to navigate the extensive volumes of radio content, it becomes crucial to establish a more precise method for differentiating between similar audio content and distinctly new sounds. This necessity arises from the inherent characteristics of historical sound recordings. Historically, due to economic constraints, there has been a concerted effort to minimize silences in sound recordings. Storage media such as vinyl and tape were costly, and in radio broadcasting, silence posed a risk of losing the audience’s attention (*Scannell 1986*). Consequently, it becomes imperative to identify changes in audio content through means other than pauses. In delving into the vast archives, the primary focus should be on segmenting sounds that represent a significant departure from the preceding audio. The question of what fundamentally distinguishes one sound from another has its theoretical origins in ancient Greek philosophy, but for our purposes, we shall restrain ourselves to more contemporary debates and perspectives.

The mid-20th century witnessed a resurgence of interest in the question of sound objects, spearheaded by thinkers and composers like Pierre Schaeffer (!Schaeffer, 1966-1967). As early as 1966, another French theoretician, Abraham Moles, made significant strides towards a concept of the aural distinction. Moles himself was also interested in radio, considering it an “art of time” (*Moles 1968*). Working on the convergence between aesthetics and information theory, Moles sought to define the experience of sound at its signal-based core. We can use his definitions today in order to distinguish between different approaches to recorded sound. His proposition was that perception of sound had several different temporalities, corresponding to different stages of autocorrelation. The algorithm above concerns what Moles would call the “time of the present”. This level of perception echoes Husserl’s first degree of retention, an awareness of the temporally coexisting (!Husserl, 1991). Yet, sound is perceived on other temporal scales as well. Moles, who was highly interested in thresholds, posed that, if the amplitude analysis of audio events is related to the limit between the “absolute and its saturation”, human perception of timbre distinguishes between small-scale differences. Thus, new sounds can be considered as events with different spectral form. But what constitutes a spectral form?

Moles made the reasonable assumption that a spectral form should exhibit some level of continuity or autocorrelation. This idea can be operationalized into a detection method that seeks out prolonged similarities in spectral activity over time. By narrowing our focus to the perceptually relevant ranges of sonic activity, we will primarily analyze the data using mel-frequency cepstral coefficients (MFCC). Widely adopted in contemporary audio processing, MFCC, in a simplified explanation, assigns weights to portions of the audio spectrum that are significant for human hearing. Given that the human auditory range typically spans from 20 to 20,000 Hz, this method involves taking the logarithm of the input to produce results that more accurately mirror human auditory perceptions. The script presented below processes the audio file and generates the MFCC across 13 spectral bands.

To interpret this image, it is possible to imagine that we tilt it 90 degrees to the side. This would result in a sequence of small power spectra, similar to the one plotted above. This method provides a means to consider the spectral development over time and thus come closer to the desirable “timbral” character of a sound. I will not diverge into the long and heated history of debating timbre. For this approach, it suffices to consider that there are today several popular methods for inquiring into the shape of sounds and how they change over time. In the following example, two such approaches are compared on the audio above. For each frame in the audio, the normalized MFCC and the spectral centroid have been calculated and plotted on the graph.

Both methods provide a visual indication of similar segments. Instead of detecting several different occurrences within the musical segments, we receive one stable period of similarity. Nevertheless, the method is still suspectable to silence. In the interview segment towards the end of the file, longer pauses between answers still render gaps which would translate into new events. As the purpose here is to grasp the more significant segment changes within hundreds of hours, such granularity can be left out. In order to achieve this, it is valuable to not only be able to measure the lasting similarity of acoustic events but also their respective difference. Since the results of the normalized MFCC utilize a greater range of values, the difference will be easier to measure. Thus, in the following application, I leave the central spectroid aside.

To put this into action, a simple code can be devised that takes the frame-by-frame MFCC value and compares it for similarity. This algorithm analyzes the MFCC values frame by frame, assessing them for similarity. When the average value across a certain number of frames remains consistent, the algorithm identifies this as a detected sound. Given the high frequency of new detections this method can generate, the algorithm is enhanced by leveraging variations in the normalized MFCC results. It compares the MFCC values of each newly detected epoch of similarity with those of the preceding one. To prioritize sonic forms extending beyond the ephemeral ‘time of the present’, the algorithm is designed to recognize similarities that persist for more than 10 frames, which approximately equates to three seconds. The code in the hermeneutic layer contains thresholds for similarity and difference. They control the amount of variation within the MFCC values which are required for annotating a new sound. There is no universally applicable setting available. Each audio data has its particularities, and for the following analysis, I’ve deduced an appropriate threshold by testing. The method used below is not more complex than that.

On this rather granular scale, it becomes apparent that the method bundles together content rather crudely. But as we scale the method up in the next section, it will provide useful discrimination applied to the hundreds of hours of radio content. When exploring the more general character of such a large set of data, I will argue for pre-processing by means of more rudimentary analysis. While there are effective pre-trained models for identifying specific sounds in a recording, applying granular object classification to radio content would be a waste of its capacity. The absolute majority of content will be either speech or music, resulting in a copious number of labels of the same type. By segmenting the audio in the manner mentioned above, it is possible to focus computational resources on the less-known aspects of the data.

This description summarizes the method of preprocessing the audio in the following study. I hope to have demonstrated the value of lingering upon this initial step of the analysis. Preprocessing might sound arbitrary, but these decisions set the frame for the very analysis ahead. And in regard to historical radio data, there really are no predefined standards. As noted, many of the choices border on the arbitrary, serving only the function of providing an effective threshold for this particular material. I therefore also encourage the reader to access the hermeneutics layer and explore the settings, and the different results they generate. The approach suggested here is intended to be easily replicable, with the aim of further improvement and updating (*Clavert, Fickers 2022*).

## The sample set and its historical significance

As the attentive listener might have noticed, the sample sound inspected above is in fact not Swedish radio, despite this being the topic of the historical analysis. As mentioned, the reason for this is a strict legal interpretation that prohibits the spread of these files. The old American samples have served as a means to grant better understanding of the method and algorithms in use. Beyond this point, however, the analysis will deal with data that is protected. Therefore, the files only contain relevant acoustic data extracted from the actual files. The legally restricted status of the data has also affected my access, limiting the sample size used in the study. In order to analyze the data, physical presence at the archive or streaming have been the only options. Since both analysis and testing are time-consuming, this limits the sample size. Direct access to the files would have allowed for a far more comprehensive study. The methods applied would nevertheless be valid for much larger data collections. Nevertheless, the results can still provide interesting indications.

The aim of this inquiry is to gain insight into the stylistic development of public service radio over time. The aesthetics of public broadcasting, and its distinction towards commercial radio, has been historically debated both within and beyond media studies. In theory, the distinction is quite clear. Commercial radio is considered economically dependent on audience numbers, whereas public service supposedly has autonomy against these matters for the benefit of other values. However, as Denis McQuail has shown, this is not always such a clear case in practice (*McQUAIL 2003*). The structure of PSB can vary widely, ranging from efforts to be more cost-conscious to scenarios where distribution networks and entire channels are sold off (*Steemers 2001*). Numerous studies provide evidence of this spectrum in real-world cases. This critical examination of PSB has predominantly focused on organizational and economic aspects (!Gloria, 2018). Given that PSB is defined largely in economic terms, it might seem logical to conceive of it as a spectrum encompassing a variety of real-world examples. Yet, this raises an intriguing question: what about the actual content of the broadcasts? Is there a discernible, non-commercial style of broadcasting that characterizes public service radio?

For this purpose, the analysis below considers a sample set of 10 randomized sample days from two years before the introduction of commercial radio; 1988 and 1991, and two years after; 1994 and 1998. The aim of the sample is to capture the weekend content flow, and thus weekdays have been omitted. While a comparative analysis of weekends and weekdays could offer fascinating insights, such exploration is beyond the scope of this article. The focus is on the two most popular PSB channels, P1 and P3, to gain an understanding of the changes occurring in conjunction with the advent of commercial radio. The sample encompasses approximately 18 hours of broadcasting per day, totaling around 1600 hours of radio content. The actual data is split into several files for each day and requires manual work to piece together and normalize. Nonetheless, with its uniquely early legal deposit regulation, the Swedish media archive offers a unique insight into the changes during this epoch. The possibility to study entire broadcasting days, instead of selection by the radio company themselves, grants a rich understanding of the global structure of broadcasting, and how it changes over time. The data can be investigated below:

## Analysis

Through the method outlined in the previous sections, the extensive audio data from each yearly set can be condensed into a collection of sample blocks. As demonstrated earlier, the audio has been analyzed for breaks in similarity sequences, from which 3-second-long samples have been extracted. Due to legal restrictions, I have extracted the MFCC values in advance. The result presents the most significant sonic sequences within the data. Notably, there is a slight shift in the number of detected occurrences for each year. Where the algorithm detected 1433 objects in the sample data from P3 in 1988, only 1030 were found a decade later. The decision has been made to keep these statistical variations, as they reflect the overall structure of the broadcasting. However, this slight imbalance should be kept in mind as we explore the distribution between the years. To better understand the nature of these sonic features, we can begin by mapping their internal relationship using complexity reduction techniques and plotting them in Euclidean space.

PCA (Principal component analysis) is a well-known tool in the digital historian’s toolbox, and its mechanics can be studied here (ref). In the plot, MFCC data has been extracted from each 3-second sample and then subsequently aligned on the axis of the PCA plot. The following code enables switching between PCA and the different scales of UMAP (Uniform Manifold Approximation and Projection) to explore the dataset. By comparing both methods, we gain better understanding of these limitations. PCA is a linear technique that transforms data into a new coordinate system where the first axes capture most of the variation in the data. While PCA is good at preserving the global structure of data, it can miss non-linear patterns or relationships among the data. PCA is also sensitive to the scaling of features and can be affected by noise or outliers. UMAP, on the other hand, is a non-linear technique that has proven more likely to preserve the local structure of the manifold. UMAP is better at separating different groups of data and capturing non-linear variations in data. UMAP is also more flexible and scalable than PCA and can handle different types of data. However, UMAP is more complex, random, and sensitive to hyperparameters such as the number of neighbors or minimum distance.

In both graphical representations, the data from P1 and P3 for the year 1988 are superimposed for comparative analysis. The visual results of P1 data points reveals a broader distribution, indicating a greater variance in the MFCC data. In contrast, the data for P3 is more densely clustered, implying a higher degree of similarity among its samples. However, when we turn our attention to the UMAP results, an interesting pattern emerges. While both channels exhibit a concentration of content in similar regions, P3 displays distinct, separate clusters, which could be indicative of unique sound types not present in P1’s broadcasts. This observation is particularly noteworthy given UMAP’s reputed efficacy in capturing local cluster formations. Interpreting these results through the lens of sonic identity, one might conjecture that P1’s sound profile is somewhat less defined or distinct compared to P3.

Granted, this is a somewhat experimental approach. There are for example many other approaches to dimension reduction available. PCA and UMAP have been chosen as two popular, but different methods. However, it’s vital to acknowledge the heuristic nature of such tools — while they can shed light on general patterns and structures in the data, they might not always capture its granular nuances. Authors like Orit Halpern and Johanna Drucker have effectively stressed the ideological depth of data visualization. My ambition with the use of two comparative methods of visualization is to further emphasize the critical hermeneutics of audio segmentation. If the previous segments demonstrated how complex the focus of audio segmentation is, this comparative analysis is intended to dispel the idea that “data and visualization were one and the same” (Drucker, 2020: 12). Following Drucker’s argument, it would be possible to go further away from the standardized methods of distance measurements, yet this would invite a whole different set of questions. Currently, I am interested in applying critical hermeneutics to primarily highlight the complexity of audio processing.

To better understand the degree to which these results actually capture significant qualities in the sound, such an unsupervised method can be complimented with a pretrained module for object recognition. Today the methods for sound classification have grown mature and advanced, offering relatively reliable classification capabilities for content categorization. Due to the availability of massive training data sets, the likes of “audioset”, the development progresses rapidly (14492245/TPC5DAIR). However, this type of classification can be quite processing-heavy and would be straining to apply to the 1600 hours of data. Thus, this exposes one of the benefits of the pre-processing in which we have already extracted a representative selection of the variation of different sounds.

In the following figure, the EfficiantAT model has been used to classify the samples. EfficiantAT is at the time of writing one of the best performing audio classification solutions around, offering a set of pretrained models with varying complexity. The model “mn20\_as\_ext” has been employed, which is one of the more granular and accurate models. To get a sense of the reliability of the results, the following graph displays the security percentage on the data from the separate highest-achieving models. As the audio files are not public, this process cannot be included in the hermeneutical layer. Therefore, the following graphs are copied from my private notebooks.

There are some eyebrow-raising identifications. The fact that the algorithm detects both an “Oink” and a “pig” may strike the reader as a rural hallucination. To my own surprise, I discovered that there were in fact instances of pig sounds in one of the sample days, linked to a reportage on farming. The same result came from exploring the enigmatic presence of a “goat” in the audio. The “snake” was harder to identify, the sound was more likely caused by a microphone noise.

Certain distinctions are also questionable in their detail. While a broad classification could encapsulate many of these under the umbrella of ‘music,’ the algorithm discerns finer variations, identifying distinct categories like choir, whistle, and opera. This nuanced differentiation underscores the predominance of speech and music in radio content, echoing the early theoretical perspectives of Rudolf Arnheim on radio (Arnheim, 1936). Arnheim noted that radio, as a medium, is confined to a limited palette of semiotic expressions, with speech and music being its most prominent components. His distinction did not anticipate the occasional representation of farm animals.

More importantly, however, the results reveal interesting discrepancies between the two approaches. Considering the variation of content types revealed by classification, it is clear that the dimension passes over certain distinctions within the sound. However, it is not that PCA is blind to differences – basing the distribution on MFCC values just allows for distinctions that are unfamiliar to more established cultural categories. Where a “chant” and a “mantra” may appear very distinct to the listener engaged in these activities, the sonic qualities are much less distinct than between two different mantras. Thus, the approach is limited in its grasp of cultural signification. However, in another sense, such a shift in perspective allows new ways of thinking about the sounds of radio. The method enables inquiry into the global sonic characteristics of the radio. From this perspective, it becomes all the more interesting to explore the data over time. In the following figures, P1 data is plotted over four sample years from the decade, respectively. To aid the visual estimation, the average pairwise distance value for each sample set is printed below.

Upon initial examination, the results from P1 appear to contradict prevailing expectations. A notable concentration of spectral variation is observed up until 1991, prior to the advent of commercial radio. More intriguingly, the data from the subsequent year reveal a noteworthy increase in diversity. This finding is particularly significant as it seems to run counter to the stated objective of establishing a more pronounced identity for P1. Internal documents from 1993 emphasize the importance of aesthetic consistency in cultivating this identity, to the extent of standardizing the logos on microphones. Given this context, the apparent shift in the opposite direction within the sample data is quite striking. Before delving into possible interpretations of this trend, let us consider the P3 results.

In contrast, P3 displays a wider distribution of sound, already from the outset. The possible cause for this could be related to the reliance of popular music on P3. Previous research has shown that popular music witnessed a decrease in timbral variation during the 80s (*Mauch, MacCallum, Levy, Leroi 2015*). This implies that the songs and artists in popular rotation during the 80s were employing a smaller and more predictable part of the frequency spectrum. However, by the end of the decade, data has displayed how frequency variation significantly increases again. Given that P3 kept a rotation mostly consisting of popular contemporary music, it would be reasonable to assume that it would reflect this increase in spectral variation over the years. Therefore, the increase in variation in the late 80s could be part of the general increase in pop musical timbre during that period. Perhaps more interesting is the development between 1991 and 1994, which strongly resembles P1. P3 also displays its most significant change between these two sample years and in the direction of increased spectral diversification. The mean pairwise distance increases from 82 to 114, a significant change.

There are at least two possible interpretations of these results. At first glance, they seem to contradict the initial goal of creating a clear and homogeneous sonic identity. P1, the flagship of SR, appears even more diverse than the music channel P3. However, when we consider both the data sets together, another explanation comes to mind. If we consider the prospect of identity construction, not from the perspective of individual channels, but as the total sum of the PSB output, there is a common element to P1 and P3; diversity. This is an intriguing idea, especially when bearing in mind the notion that commercial radio is typically demarcated by sonic similarity. It is tempting to consider that PSB might have sought to create a general identity of variation. To further explore this conclusion, more data would be needed, especially from commercial broadcasting in Sweden during this time. There are, however, a few previous research attempts that have raised the issue. Notably, media scholar Carin Åberg conducted a pioneering content analysis of Swedish commercial radio in the 1990s. Her results pointed towards “an extremely repetitive structure” (*Åberg 1999*). In her study, only one sample day was analyzed, but the tentative results suggested that perhaps, for studying the structure of commercial radio at the time, no more was needed. Very similar segments and short top lists were repeated with short intervals. Assuming that the new commercial competition introduced on the Swedish stage in 1993 primarily suffered from this high level of predictability, higher degrees of variation would be a valid approach to identification. If this could be a partial explanation for the tendency in the data set, it is all the more curious that the results return back to their previous values just three years later. This suggests that if it was an intentional strategy at the time, it quickly lost its impact.

Due to the status of these files, the reader will regrettably not be able to examine the acoustic implications of this change. During my research, I have been able to closely listen to specific sound segments from each cluster, and the experience has tended to align with the plotted results, but only in the sense of spectral character. The perceptual experience of different spectral ranges is not reducible to the entirety of the aural experience, but we can certainly perceive the significant familiarity in the sound. Nevertheless, to give further credence to these results, we can proceed by comparison with the object classification method. In the following plot, the number of identified unique sounds from each data set has been measured. Since plotting the results on the dimension reduction map did not render any significant insight, the results are here simply calculated, in order to compare with the spectral distribution displayed above.

Perhaps surprisingly, the results do overlap with the more large-scale tendency in the dimension reduction algorithm. Though differing in amount, the general direction of change mirrors the spectral diversity rather clearly. Where P1 displays a diminishing variation of sound between the first two sample years, both channels exhibit a noticeable increase towards the third sample set, which finally declines during the year. Despite the unexpected nature of this alignment, the results are produced by the same means which was scrutinized with close listening in the previous example and should therefore carry certain statistical insight into the larger scope of broadcasted content. However, we are at this point progressing towards a scale level which is difficult to verify by human listening. The percentual increase of overall sound type variation throughout a day of broadcasting might be experienced in its statistical specificity but might instead give the listener an overall experience of less sonic consistency. Though regarded from a quantitative perspective, there remains an inherent contradiction to the project of creating distinctiveness through change. Nevertheless, these results give credence to the suggestions that SR might have been trying to generate broadcasting identity through sonic variation.

There are also specific historical indications that support this reading. It can only be briefly mentioned within this scope, but SR’s aim for sonic variation might be regarded in the context of broader attention to content diversity at the time. In the spring of 1993, the words “relative entropy” suddenly covered the pages of the Swedish press. Swedish television had won the prize for the highest entropy. This enigmatic concept was coined by telecommunication professors Jacob Wakshlag and William Adams in 1985, but received its popularization through the “Quality Assessment of Broadcast Programming” (QABP) report (*Hellman 2001*). QABP was a global research project instigated at the University of Tokyo in the early 1990s. Prompted by worries about the increasingly free-market oriented structure of mass media, researchers set out to determine a definitive measurement of “quality” for broadcasting to navigate towards. In a somewhat surprising twist, however, the answer decided upon turned out to be related to quantity. Mass media ought to provide the highest variation in content possible. In a large-scale study comparing broadcasting from 26 countries, QABP measured the quantitative variation of content on individual channels, and Swedish public service television received the best score.

This might have influenced broadcasting production. In the first place, it granted the question of entropy, variation, and unpredictability a renewed status in media production in Sweden. Secondly, because Swedish radio also had been evaluated and received a much lower score than television, it might have spurred producers in the direction of all sorts of diversity. To fully explore the correlation between these events is a matter I develop in my dissertation, and which far exceeds the scope of this text. Yet, I hope that the correlation between the global development in both object detection and dimension reduction might convince the reader that the tendency carries some value. And if so, we are also forced to consider the apparent brevity of this trend.

## Summary

This experiment has taken place on the threshold. On the threshold between listening and counting. On the threshold between differences in sound. Finally, it also plays out on the threshold to the scientifically valid. The sample sizes have been too small to actually confirm the findings on a solid base, despite strong indications. I hope that the reader still can see the value in the results. The analysis of audio data from annual sample sets reveals a shift in the number of detected occurrences for each year, with varying sonic sequences reflecting the overall structure of the broadcasting. Using PCA and UMAP, we observe differences in the MFCC data and clustering between P1 and P3, suggesting distinct sonic identities. There are several contextual factors related to the changes in the media environment that could be considered as a cause for this development, yet the exploration of causality lies beyond the confines of this article.

However, beyond these historical endeavors, my aim has also been to contribute to the understanding of historical radio data. By suggesting relatively simple methods for preprocessing and segmentation, the sonically curious reader might find inspiration for future improvements in working with cultural audio data. Furthermore, I want to stress the methodological consequences of combining dimensionality reduction and object classification. The initial observation in the analysis suggested an incongruity between untrained dimensionality reduction techniques for spectral distribution and pretrained object classification models. Yet, from as we consider the large-scale trends in the data set, certain similarities between the two methods become clear, revealing common trends in development. Such an observation points to the benefits of complementary usage of the two methods. It shows the value of allowing pretrained and untrained algorithms to be experimentally combined. Each approach allows for different insights into the local and fine-grained aspects of audio data, but together they may be utilized to test the validity of the results. It is only by combining and testing out the great variety of approaches to audio data processing that we can begin navigating the radio archive.

# Bibliography

Åberg, C.. The Sounds of Radio : On Radio as an Auditive Means of Communication. Dept. of Journalism, Media and Communication [Institutionen för journalistik, medier och kommunikation - JMK], Univ..

Bravo Sanchez, F. J., et al.. “Bioacoustic Classification of Avian Calls from Raw Sound Waveforms with an Open-source Deep Learning Architecture”. Scientific Reports, vol. 11, no. 1, p. 15733, https://doi.org/10.1038/s41598-021-95076-6.

Bösch, F., et al.. “Roundtable. Writing (media) History in the Age of Audio-visual and Digital Media”. Journal of Modern European History, vol. 10, no. 1, pp. 98–116, https://doi.org/10.17104/1611-8944\_2012\_1\_98.

Clavert, F.and A. Fickers. “Publishing digital history scholarship in the era of updatism”. Journal of digital history, no. jdh002, https://journalofdigitalhistory.org/en/article/m7DWqDjY3hoV.

Clement, T. E.. “When Texts of Study Are Audio Files”. A New Companion to Digital Humanities, John Wiley & Sons, Ltd, pp. 348–57, https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118680605.ch24.

Eco, U.. A Theory of Semiotics. Indiana University Press, https://www.jstor.org/stable/j.ctt16xwcfd.

Ernst, W.and J. Parikka. Digital Memory and the Archive. University of Minnesota Press.

Have, I.and K. Enevoldsen. “From Close Listening to Distant Listening: Developing Tools for Speech-music Discrimination of Danish Music Radio”. Digital Humanities Quarterly, vol. 015, no. 1.

Hellman, H.. “Diversity - an End in Itself?developing a Multi-measure Methodology of Television Programme Variety Studies”. European Journal of Communication - EUR J COMMUN, vol. 16, pp. 181–208, https://doi.org/10.1177/0267323101016002003.

From Stationality to Radio Aesthetics: Investigations on Radiophonic Sounds. From Stationality to Radio Aesthetics: Investigations on Radiophonic Sounds. https://doi.org/10.7551/mitpress/9975.003.0040.

Iasa-phonographic-bulletin-13.pdf. Iasa-phonographic-bulletin-13.pdf. https://www.iasa-web.org/sites/default/files/bulletin/iasa-phonographic-bulletin-13.pdf.

Papers with Code - Audioset Benchmark (audio Classification). Papers with Code - Audioset Benchmark (audio Classification). https://paperswithcode.com/sota/audio-classification-on-audioset.

Kittler, F. A.. Gramophone, Film, Typewriter. Translated by Michael Wutz and Geoffrey Winthrop-Young, Stanford University Press.

Kittler, F.and G. Winthrop-Young. “Real Time Analysis, Time Axis Manipulation”. Cultural Politics, vol. 13, no. 1, pp. 1–8, https://doi.org/10.1215/17432197-3755144.

Mauch, M., et al.. “The Evolution of Popular Music: USA 1960–2010”. Royal Society Open Science, vol. 2, no. 5, p. 150081, https://doi.org/10.1098/rsos.150081.

McQUAIL, D.. “Public Service Broadcasting: Both Free and Accountable”. Javnost - the Public, vol. 10, no. 3, pp. 13–28, https://doi.org/10.1080/13183222.2003.11008832.

Moles, A. A.. Information Theory and Esthetic Perception. Univ. of Illinois Press.

Scannell, P.. Radio Times: The Temporal Arrangements of Broadcasting in the Modern World. https://eric.ed.gov/?id=ED294542.

Steemers, J.. “In Search of a Third Way: Balancing Public Purpose and Commerce in German and British Public Service Broadcasting”. Canadian Journal of Communication, vol. 26, https://doi.org/10.22230/cjc.2001v26n1a1196.

White, H.. “Historiography and Historiophoty”. The American Historical Review, vol. 93, no. 5, pp. 1193–99, https://doi.org/10.2307/1873534.