Utilisation of Audio Mining Technologies for Researching Public Communication on Multimedia Platforms

Eble, Michael; Stein, Daniel

Erstveröffentlichung / Primary Publication
Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:
Eble, Michael ; Stein, Daniel: Utilisation of Audio Mining Technologies for Researching Public Communication on Multimedia Platforms. In: Maireder, Axel (Ed.) ; Ausserhofer, Julian (Ed.) ; Schumann, Christina (Ed.) ; Taddicken, Monika(Ed.): *Digitale Methoden in der Kommunikationswissenschaft*. Berlin, 2015 (Digital Communication Research 2). - ISBN 978-3-945681-02-2, pp. 329-345. URN: https://doi.org/10.17174/dcr.v2.14

Digital
Communication
Research.de

Suggested Citation: Eble, M., & Stein, D. (2015). Utilisation of Audio Mining Technologies for Researching Public Communication on Multimedia Platforms. In A. Maireder, J. Ausserhofer, C. Schumann, & M. Taddicken (Eds.), *Digitale Methoden in der Kommunikationswissenschaft* (S. 329-345). doi: 10.17174/dcr.v2.14

Abstract: The number and volume of spoken language corpora which are generally available for research purposes increase significantly. That is due to the wide adoption of audio-visual communication on news websites and social web platforms. The respective messages that are published by professional and individual communicators are subject to online content analysis. To date, such analyses strongly rely on manually operated processes which come along with a huge effort for transcribing spoken language corpora into textual content. Hence, challenges like the ever increasing volume, velocity and variety of multimedia content need to be faced. Audio Mining technologies are capable of reducing the effort for turning speech into text significantly. Using these technologies via application programming interfaces (APIs), it is demonstrated how a hybrid approach enables researchers to reduce the time that is needed for analysing news content by an order of magnitude.

DOI 10.17174/dcr.v2.14

Michael Eble & Daniel Stein

Utilisation of Audio Mining Technologies for Researching Public Communication on Multimedia Platforms

1 Public communication on online multimedia platforms as research subject

Everyday public communication fulfils several functions regarding transparency, validation and orientation for modern societies (Neidhardt, 1994; Donges & Imhof, 2005). Audio-visual media content has become an essential part of public communication, and it takes place on Internet platforms for two reasons: First, such content is constantly being produced by professional publicists, editors, journalists etc. for multimedia platforms like BBC.co.uk. Secondly, it is also being used in professionally motivated public communication as well as individual communication, e.g., on social web platforms like YouTube. The production and consumption of audio-visual media content enable online-based publics for certain topics like elections or sport events (van Eimeren & Frees 2012, p. 371). In addition to that, social web platforms have been established as infrastructures for networked publics and follow-up communication through strongly interlinked audio-visual media content (Schmidt, 2011 & 2013). A current example for this is the uptake of applications and formats in the area of "Social TV" and "Second Screen" (Eble, 2013b).

Overall, the number and volume of spoken language data which are generally available for research purposes in the field of public communication increase significantly. The *growing importance of audio-visual messages* for public communication can be tracked down to three reasons: First, editorial staff at online media organisations increase their production and publication of audio-visualcontent. Secondly, media organisations digitize and open up their archived audio-visual content in order to attract more users to their content and brand. Thirdly, user-generated content is no longer restricted to text but includes more and more video content instead.

Public communication that takes place on Internet platforms and makes use of audio-visual messages is subject to *online content analysis*. However, several challenges need to be addressed in order to successfully carry out such analyses of multimedia content. Given that context, the paper at hand sets out to contrast methodical challenges and approaches from the field of automatic speech analysis (section 2). Furthermore, it aims for strengthening the bridge between computer science and communication studies by giving an overview on the utilisation of Audio Mining technologies for analysing public communication on multimedia platforms (section 3). A hybrid approach is emphasised that takes into the account the advantages arising from collaborative human and computational analysis (section 4). The authors are convinced that such an approach is inevitable in order to cope with future challenges in online communication research.

2 Methodical challenges in the analysis of audio-visual media content

Content analysis is an essential and well-known research method within media and communication science (Wirth, 2001). According to Früh (2007, p. 27) it is understood as an empirical, non-reactive process for the "systematic, intersubjectively comprehensible description of content and formal characteristics of messages" that enables the quantitative investigation of messages.[1]

1 The quote has been translated by the authors from German as given in the source to the English version at hand.

Thus, quantitative content analysis is about generalised statements that are based on the systematic study of amounts of text. Therefore, content analysis sets out to identify certain patterns using a code book that is based on theoretical frameworks and individual research questions (Früh, 2007, p. 39). In addition to a formal descriptive analysis, a diagnostic approach is possible: One can infer from parameter values of the content to the communicator, the recipient and the communication situation, for example (Merten, 1995, p. 23; Früh, 2007, p. 44). In general, content analysis is characterized by a high degree of flexibility, as it can be applied on text, image, audio and video at any time, arbitrarily often and with varying research questions (Früh, 2007, p. 39). The method is mainly applied to content that is produced and published by public media institutions (Rössler, 2010, p. 34), which is due to the discipline's research tradition as well as its concentration on media services that are most relevant for society (Zeller & Wolling, 2010, p. 143).

Meanwhile, online content analysis has been well-established in order to study the structures and messages of content published by institutional online media (Welker & Wünsch, 2010). Rössler and Wirth (2001, p. 284) propose a typology for online content analyses that includes 1) a content-centred approach and 2) a user-centred approach. The first approach consists of area or category analyses (e.g. genre, format, groups of organisations) as well as focus analyses (e.g. events, issues, persons). In the second approach, they distinguish between publicity analyses (e.g. on the basis of traffic or other usage metrics) and selectivity analyses (e.g. according to user behaviour). A combination of the content-centred and the user-centred approach can increase the insights gained through this method (Merten, 1995, p. 119; Zeller & Wolling, 2010).

Online content analysis needs to deal with heterogeneous media services and interlinked multimedia messages: Due to current and upcoming technical and editorial patterns of news production, distribution and usage, the respective data is distributed among various sources. That is, research data are characterised not only by increasing volume, but also by velocity and variety of online content. With respect to such challenges, Schweitzer (2010, p. 45) sees

2 The quote has been translated by the authors from German as given in the source to the English version at hand.

the need for intensive discussions on specific recommendations for future research practice. This also seems to be strongly required according to Rössler (2010, p. 41), who argues that expertise in the field of online content analysis could become a "key qualification for empirical research in all social science disciplines".[2]

With these considerations in mind and with respect to audio-visual content, challenges of online content analysis can be described in more detail as follows (Eble, 2013a; Eble, Ziegele, & Jürgens, 2014):

- The *definition of research samples* can be challenging or almost impossible in terms of the sample's determinability and definition of the selection unit.

- The subsequent *collection of audio-visual data* is challenging especially in terms of its (from time to time hardly predictable) dynamic and constant flow (e.g. live streams) as well as its volume. That requires appropriate automatic software agents (crawler).

- Having collected the data that is needed for one's research question leads to the need for *archival and transfer of data* which comes along with challenges like storing raw (i.e. unstructured) and analysed (i.e. structured) data in storage systems most suitable for the respective purpose (e.g. HDFS, NoSQL or SQL databases)[3]. Storing the initial raw data apart from the results of its processing puts the researcher in the position to re-analyse the messages based on future research questions.

- Finally, for the *analysis of data*, issues like avoiding live coding, encoding and re-coding, ensuring intra-coder and inter-coder reliability as well intersubjective confirmability, etc. have to be tackled (McMillan, 2000, p. 93; Rössler & Wirth, 2001, p. 296; Weare & Lin, 2000, p. 287). Furthermore, researchers have to deal with legal and ethical issues in terms of data privacy and personal rights (Eble et al., 2014).

3 HDFS: Hadoop Distributed File System; NoSQL: Not only SQL; SQL: Structured Query Language.

Given that context, the following section focuses on the automatic transformation of spoken content into textual content via audio mining technologies in order to enable researchers to carry out the actual content analysis much faster. Thus, the content within audio-visual media, e.g., the transcript of the interview with a politician derived from speech, becomes conveniently accessible for manual coding by human annotators which use tools like SPSS or QDA Miner and apply a specific coding scheme. That is, the paper at hand primarily addresses challenges occurring with data analysis. In addition, some insights concering challenges of storing and transferring data are provided.

3 Supporting content analysis through Audio Mining technologies

In order to illustrate challenges that arise in communication science research when working with audio-visual media content, and to elaborate on approaches that computer science could offer, we make use of the following (conceptual) use case scenario:

> Jessica Müller, PhD student, is investigating the (presumed) change of journalistic reporting on renewable energies and their positive and negative effects as a focus analysis on the topic "renewable energy". This content analysis is focusing on four major broadcast news shows in the time frame of January till December 2014, working with ARD, ZDF, RTL and SAT1 which have the highest media coverage (selected in terms of publicity aspects). The creation of appropriate categories has already been conducted, and the code book is present. In order to move on, Jessica now needs access to transcribed text of the spoken content of all the video files.

Based on the author's experience a manual transcription of these broadcast news shows seems impractical. Assuming a real-time factor of seven for the manual transcription of words (in a very basic manner such as: words, interjections and speaker changes; more elaborate schemes such as, e.g., GAT2 (Selting et al., 2009) would take considerably more time) in an audio signal means that a daily broadcast of a quarter of an hour over the course of a year would result in roughly eighty working days. It thus seems reasonable to employ automatic speech analysis (ASA) as a pre-processing step, at the very least to identify the videos relevant for further, manual post-processing. In direct comparison to other media processing techniques, ASA is comparatively old and well-established. Many new

systems have adopted a deep neural network (DNN) paradigm (Hinton, Deng, Yu, Dahl, Mohamed, Jaitly, & others, 2012), and it seems that these machine learning architectures that consist of multiple non-linear transformations are quite capable in handling speech input. With modern computer power, they are also real-time capable and outperform former approaches with more rigid architectures quite consistently.[4]

> Jessica processed the videos with automatic speech analysis. On a computer with eight processor cores, 365 hours of video (assuming a proper audio/video compression and a resolution of 512x288, this equals to roughly 150 GB) in less than two days. She is searching for the first-best video that matches her search query "renewable energy" and checks the transcript for quality. Especially those proportions of the video where the anchor is speaking are quite intelligible – once she adjusts to the fact that the transcripts feature no punctuation marks. In an interview with a local resident that has quite a thick accent, the quality drops significantly. Jessica also notices that some words she is looking for are transcribed rather odd or even wrong, e.g., "null Energie Haus" rather than "Nullenergiehaus"; and "Off Shoah Wind Parks" rather than "Offshore-Windparks".

Especially for broadcast news, ASR is considered to be well-established. For planned speech with clean audio quality (i.e., sufficient sample rate with no or little background noise) and a known speech domain (news, sport, tourism) it shows very good results (Jurafsky & Martin, 2000). However, it is also true that deviations[5] from this ideal scenario are still in the focus of current research (Goldwater et al., 2010). These deviations include: spontaneous speech, together with hesitations, interjections and incomplete sentences; background noise (and especially music, which has strong similarities to speech characteristics because of its harmonic nature); multiple, simultaneous speakers; and pronunciation variants due to speech disorders, accentuated speech or loan words (German: "getwittert", "unfriended", "bromance"). The underlying reason for some of these errors is the internal process of the ASR system, which is typically built around recognizing phonemes before mapping phoneme chains to actual words. Unknown

4 For the systems used for our own research projects, we witness an error reduction of about one-third.

5 The term "deviation" is meant to be mathematical here and indicates a statistical deviation with respect to the models derived from ideal data.

words (i.e., never seen before during the ASR training phase) and neologisms such as "Nullenergiehaus" are thus as problematic as words that might have been encountered in previous texts but are pronounced in a uncommon way (e.g., lean words such as "Offshore-Windparks" or local pronunciation variants).

> In order to quantify the quality of the transcripts, and to decide whether they are appropriate for her studies in spite of the errors, Jessica corrects the transcripts of a bunch of videos by hand and computes the word error rate. For the selected videos, she obtains an error rate of 15 percent.

The common quality criterion for automatically derived transcripts is the word error rate (WER), which is the number of transcription errors weighted over the total number of reference words. Transcription errors include insertions, deletions and substitutions (also called "Levenshtein distance" or "edit distance"; Levenshtein, 1966). In an ideal scenario, WERs that are below ten percent are often witnessed. Studies suggest that a WER of up to 25 percent is still perceived to be intelligible (or at least understandable) by humans, and that WERs of up to 40 percent still render the transcripts to be usable by further text mining algorithms (Munteanu et al., 2006). ASR systems also typically mark words with a confidence measure, which indicated how sure the system was during the recognition. They can thus be employed to discard those words that are considered to be most error-prone.

> Jessica feels that the WER of 15 percent is reasonable to obtain a first impression on the data, but for reliable conclusions for her analysis she deems this error rate to be too high. She decides to go for a hybrid approach, by using data mining techniques in order to identify those parts of the data that are important for her research, and to manually correct the transcripts of these parts in a second run.

In general, technologies from the field of data mining have the task to derive new, i.e., not already explicitly stated, patterns and information from (largely arbitrary) data. The algorithms and techniques employed draw their theoretical background from the fields of artificial intelligence (machine learning) (Han & Kamber, 2006) and are typically statistically-based.[6] Depending on the source and

6 I.e., they do not consist of manually defined rules but are based on statistical data-observation.

modality of the data, the techniques are also grouped into "text mining", "web mining" (here, this is primarily the analysis of web-based information such as interlinking of websites, server statistics, and user behaviour) and "social mining" (activity within social web platforms). Following an established definition (Fayyad, Piatetsky-Shapiro, Smyth, 1996), aspects of data mining are (a) grouping of similar content, (b) classification, (c) regression analysis, (d) association analysis, (e) outlier detection and (f) summary.

> Jessica wants to employ key word extraction on her documents in order to find those excerpts that are the most relevant to her term "renewable energy". She computes the 40-best (i.e., the 40 most probable candidates based on the underlying statistical model) key words automatically, but multiplies these values by the confidence of the ASR transcript output, in order not to overestimate possibly erroneously recognized words.[7]

A keyword is part of the document itself and highlights words that seem to disambiguate the document in comparison to other documents; thus, a keyword is not a stand-alone description of a document's content but rather a highlighted feature in comparison to a set of other given documents. A well-established technique is the term frequency inverse document frequency (TF-IDF)[8] measure, i.e., a word is considered the most relevant if it appears considerably often in the given document but on the same time does not appear in many other documents. TF-IDFs also show good performance for ASR transcripts (Schneider & Tschöpel, 2010). For Jessica's query, the key word actually is composed of two words ("renewable energy"), which is typically coped by modern algorithms. Especially in morphologically rich languages such as German, the words are often reduced to the stem form by automatic means and compound words can be split so that the

7 The TF-IDF algorithm used for keywords will be very keen on words that seem to be unique for a specific document/transcript and thus give it high rankings; thus, especially for words where the speech recognized did not assign high probabilities, i.e., it was unsure if this word was recognized correctly, the TF-IDF score needs to be adjusted in order to not over-emphasize wrong transcript parts.

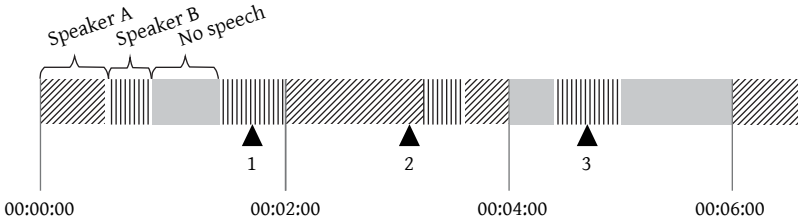
8 The term "frequency" appears twice in this name since there are two frequencies compared against each other: the frequency of the words within the document versus the frequency within all reference documents.

algorithms do not treat similar words such as “Energie”, “Energien” and “Solarenergie” as completely different entries.[9]

> Employing the key word technique, Jessica now samples a smaller corpus of 200 videos which contain the key word “renewable energy” with a reasonable ASR confidence. This is quite a step ahead from the originally 1.400 videos in her collection. However, those videos are still quite long and poly-thematic, and large proportions are not relevant for her analysis. Since she knows that each topic is introduced by the anchor, she decides to segment the videos based on speaker changes. After the segmentation, Jessica marks a few proportions where an anchor speaking (there are roughly 5-10 anchors per broadcast show), which results in the new segments to be mostly mono-thematic and much smaller than before.

Segmenting audio data based on speaker changes is also called “diarization”. This does not require known speakers; instead, the segments are assigned to alpha-numerical values such as speakerA-speakerB-speakerA-speakerC. This is called speaker identification (SID) if the individuals are known in advance. One state-of-the-art approach usable for both tasks is the i-vector paradigm (Dehak et

Figure 1: Exemplary result of a segmentation and analysis process performed by Audio Mining



Example for segment no. 1
Speaker B: “Renewable energy is one of the most important topics in the bread and pretzel industry.”

9 E.g., by Snowball stemming as part of the Lucene project (lucene.apache.org)

10 C.f. http://ivectorchallenge.nist.gov

al., 2011). Here, every segment is projected into a sparse dimensional vector space where different speakers have a large distance to each other, while different segments of the same speaker are located more closely. In international benchmark activities where different research groups compete against each other on the same data, i-vectors show very good performance;[10] moreover, they have the benefit that the sparse feature space makes it quite convenient to store after the first processing, thus building up characteristics for each segment that can be used fast for later retrieval, e.g., when the importance of a speaker becomes apparent only after the first processing.

> The segments that Jessica marked as relevant now have a total duration of 300 minutes. Jessica has them corrected manually based on the existing transcripts, which she is relieved to find to be way faster than transcribing from scratch, i.e., about double real-time. For a manually confirmed transcription of the data she is interested in, she now needs roughly 2 days instead of 80. Additionally, her computer was busy for 2 days whilst the ASR was running, and she spent another working day for the speaker diarization and recognition. With the weekend approaching soon, she decides to further improve her selection before she carries on with the entries in her code book. One thing that she is interested in is an estimation of how much segments are similar to the ones she has identified so far, i.e., documents that discuss renewable energy without mentioning this particular term.

Deriving similarity measures between different documents is used for recommendation ("users looking at this document are also interested in...") or classification of text (e.g., to identify content not suitable for minors based on the language; Bardeli et al., 2013). Similarity measures can be based on quite a large selection of features, but often also employ a bag-of-words approach, which takes the vocabulary of the document into account.[11] The techniques working with these similarity measures thus can identify documents that belong to the same topic but use different vocabulary, which can be extremely helpful when the meaning of specific terms change or if they are actually invented and established over the course of time.

11 A bag-of-word is a vector of integer numbers which counts word appearances given an overall vocabulary but ignores the position of the words within the document (so, typically, many entries will consist of zeros whenever those particular words do not appear within the document). It is an essential feature in many text mining applications.

The techniques depicted here are only a small excerpt of the tools that data mining can offer. For example, rather than looking at text transcripts only, investigating phenomena on paralinguistic aspects such as prosody, emotion, the structure of the speaker changes within heated discussions or the linking strategy of a user can be of interest. Audio fingerprinting approaches, which identify known data that has been used in other contexts, can generate valuable information, e.g., by identifying when snippets of an interview are referenced in subsequent shows (Bardeli et al., 2012).

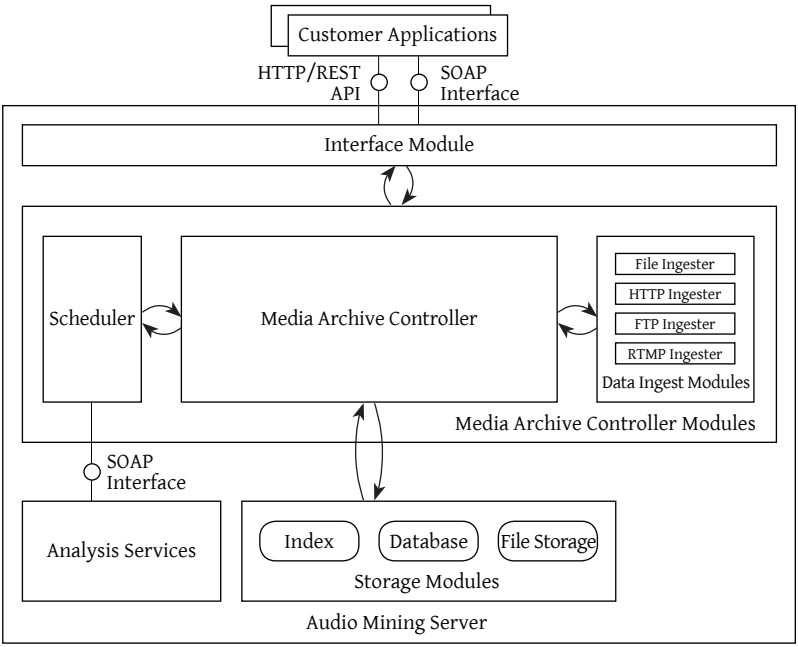
For most of these techniques, open-source solutions exist (e.g. for transcribing and maintaining: ELAN, EXMARaLDA; for natural language processing: Mallet, NLTK, Weka). However, while the algorithms itself come usually for free, they have to be trained in advance with appropriate training data (e.g. for German-language news content) before they become fully usable.

4 Transfer: System architecture for Audio Mining technologies

With respect to the use case scenarios and technologies as described above, a system architecture for researching public communication on multimedia platform can be described as follows (see FIcontent wiki, 2013): *Storage Modules* take care of capturing all the video data to file storage (e.g. as MPEG-7) and result from processing in indices and databases (e.g. XML, Solr), while *Media Archive Controller Modules* deal with all processes that need to be carried out for ingesting audio-visual media content coming from several sources and scheduling its analysis (e.g. downloading video data, processing data, exporting transcripts). A component called "*Analysis Services*" performs the actual automatic speech analysis that is relevant for research purposes and has been described above. Finally, the *Interface Module* is designed to enable application development on top of the Audio Mining system. That is, a browser-based web interface enables its users to view transcripts of spoken video content. Figure 2 (next page) comprises the modules described beforehand.

Since systems are often specifically designed for certain research purposes, there is a need for the flexible implementation of tools for automatic speech analysis via open and documented application programming interfaces (APIs). For the

Figure 2: Architecture of Audio Mining system



Source: FIcontent wiki, 2013

system described above, the API documentation[12] and a developers' guide can be found online. Therefore, using an API to automatic speech analysis for German-language content in a specific study design could enable researchers to implement scenarios as the two following (Eble, 2013a, p. 238.): The *combination of focus analyses and publicity analyses on multimedia platforms* sets out to investigate certain topics of public interest (e.g. "renewable energies") both regarding their initial messages on

12 API documentation: http://wiki.mediafi.org/doku.php/ficontent.socialtv.enabler.audiomining

13 Developers' guide: http://wiki.mediafi.org/doku.php/ficontent.socialtv.enabler.audiomining.developerguide

news sites like interviews with politicians on BBC.co.uk and their follow-up communication on social web platforms like YouTube and Facebook. Due to recent developments like Social TV, the linkages between public communication and interpersonal communication become more important. The *(near) real-time capturing of audio-visual news streams* puts researchers in the position to perform content analysis of unforeseeable news events. Due to increasing, dynamic streams of data relevant for news events and its diffusion and follow-up communication, the need for continuous data collection and according capturing systems increases, too. Otherwise, the short periods of diffusion and follow-up communication tend to prevent a reasonable retrospective study (see De Fleur, 1987, p. 125). Thus, a continuous and automatic data collection for a standardized set of parameters of a certain number of wide-reaching news sites would be a promising approach, i.e. for (comparative) re-analysis. Note that issues regarding data privacy and protection need to be addressed in such online research. Current algorithms are capable of ensuring data protection during as well as after the collection and processing of multimedia data.

5 Conclusion and outlook

The paper contrasts methodical challenges and approaches from the field of automatic speech analysis. Furthermore, it provides information about Audio Mining technologies and their use for analysing public communication on multimedia platforms in order to strengthen the bridge between computer science and communication studies. It is demonstrated how the specific use case scenario of a focus analysis on the topic "renewable energies" can be supported by automatic approaches: Speaker segmentation and identification enable researchers to efficiently split audio-visual corpora into individual parts that can be skimmed quickly. Following that, speech analysis allows for turning spoken words into text so that the content becomes accessible for content analyses. It has also been shown how a concrete system implementation of Audio Mining can look like. The FIcontent API enables developing applications for research purposes in order to carry out content analysis on multimedia platforms and to automatically handle heterogeneous multimedia data.

In further research, at least three topics need attention: First, the combination of automatic processing and human analysis needs to be improved in terms

of tools that are easy to use. To date, several data sets are available for training, like those provided by the *Linguistic Data Consortium* (LDC; www.ldc.upenn.edu) or the *European Language Resources Association* (ELRA; www.elra.info). Secondly, algorithms and their implementations need further improvement in order to handle the ever increasing amounts of audio-visual news content in (near) real-time. Thirdly, collaboration between communication studies and computer science needs to be strengthened in order to deeply understand each other's concepts, research questions and approaches.

Dr. Michael Eble is Consultant at mm1 Consulting & Management PartG in Stuttgart

Dr. Daniel Stein is Senior Research Scientist at the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS in Sankt Augustin

References

Bardeli, R., Becker, S., Bergholz, A., Kolb, I., Korte, H., Maaskant, W., Paaß, G., Schneider, D. Stein, D., & Tschöpel, P. (2013). *Studie zum technischen Jugendmedienschutz: Möglichkeiten und Grenzen von Verfahren zur Detektion jugendschutzrelevanter Web-Inhalte*. Sankt Augustin: Fraunhofer IAIS.

Bardeli, R., Schwenninger, J., & Stein, D. (2012). Audio Fingerprinting for Media Synchronisation and Duplicate Detection. Media Synchronisation Workshop, Berlin, Germany, October 2012.

De Fleur, M. L. (1987). The Growth and Decline of Research on the Diffusion of the News, 1945-1985. *Communication Research, 14*(1), 109-130. doi: 10.1177/009365087014001006

Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., & Ouellet, P. (2011). Front-End Factor Analysis for Speaker Verification. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(4), 788-798.

Donges, P., & Imhof, K. (2005). Öffentlichkeit im Wandel. In H. Bonfadelli, O. Jarren, & G. Siegert (Eds.), *Einführung in die Publizistikwissenschaft* (p. 147-175). Bern, Stuttgart, Vienna: Haupt.

Eble, M. (2013a). *Medienmarken im Social Web: Wettbewerbsstrategien und Leistungsindikatoren von Online-Medien aus medienökonomischer Perspektive*. Berlin: LIT.

Eble, M. (2013b). Social TV, Second Screen und vernetzte Öffentlichkeiten: Kommunikationswissenschaftliche Perspektiven auf Schnittstellen zwischen Fernsehen und Social Web. In U. Breitenborn, G. Frey-Vor, & C. Schurig (Eds.), *Medienumbrüche im Rundfunk seit 1950 – Jahrbuch Medien und Geschichte 2013* (p. 73-89). Cologne: Herbert von Halem.

Eble, M., Ziegele, M., & Jürgens, P. (2014). Forschung in geschlossenen Plattformen des Social Web. In: M. Welker, M. Taddicken, J. Schmidt, & N. Jackob (Eds.), *Handbuch Online-Forschung. Sozialwissenschaftliche Datengewinnung und -auswertung in digitalen Netzen* (p. 128-154). Cologne: Herbert von Halem.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17*(3), 37. doi: 10.1609/aimag.v17i3.1230

FIcontent wiki (2013). Audio Mining Specific Enabler for Social connected TV platform. Retrieved from http://wiki.mediafi.org/doku.php/ficontent.socialtv.enabler.audiomining

Früh, W. (2007). *Inhaltsanalyse: Theorie und Praxis*. Konstanz: UVK.

Goldwater, S., Jurafsky, D., & Manning, C. D. (2010). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication, 52*(3), 181-200. doi: 10.1016/j.specom.2009.10.001

Han, J., & Kamber, M. (2006). *Data mining: concepts and techniques*. San Francisco, USA: Morgan Kaufmann.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29*(6), 82-97.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing an introduction to natural language processing, computational linguistics, and speech*. New Jersey: Pearson.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*, 707-710.

McMillan, S. J. (2000). The Microscope and the moving Target: The Challenge of applying Content Analysis to the World Wide Web. *Journalism and Mass Communication Quarterly, 77*(1), 80-98. doi: 10.1177/107769900007700107

Merten, K. (1995). *Inhaltsanalyse. Einführung in Theorie, Methode und Praxis* (2. Auflage). Opladen: Westdeutscher Verlag.

Munteanu, C., Baecker, R., Penn, G., Toms, E., & James, D. (2006). The effect of speech recognition accuracy rates on the usefulness and usability of webcast archives. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 493-502. doi: 10.1145/1124772.1124848

Neidhardt, F. (1994). Öffentlichkeit, öffentliche Meinung, soziale Bewegungen. In F. Neidhardt (Eds.), *Öffentlichkeit, öffentliche Meinung, soziale Bewegungen* (p. 7-41). Wiesbaden: Westdeutscher Verlag.

Rössler, P. (2010). Das Medium ist nicht die Botschaft. In M. Welker & C. Wünsch (Hrsg.), *Die Online-Inhaltsanalyse. Forschungsobjekt Internet* (S. 31-43). Köln: Herbert von Halem.

Schmidt, J. (2011). *Das neue Netz :Merkmale, Praktiken und Folgen des Web 2.0* (2. Auflage). Konstanz: UVK.

Schmidt, J. (2013). Onlinebasierte Öffentlichkeiten: Praktiken, Arenen und Strukturen. In C. Fraas, P. Meier, & C. Pentzold (Eds.), *Online-Diskurse. Theorien und Methoden transmedialer Online-Diskursforschung* (p. 35-56). Köln: Herbert von Halem.

Schneider, D., & Tschöpel, S. (2010). A lightweight keyword and tag-cloud retrieval algorithm for automatic speech recognition transcripts. *11th Annual Conference of the International Speech Communication Association*, Imakuhari, Japan, September 2010, 1277-1280. Abgerufen von http://publica.fraunhofer.de/dokumente/N-198426.html

Selting, M., et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion, 10*, 353-402.

Van Eimeren, B., & Frees, B. (2012). Ergebnisse der ARD/ZDF-Onlinestudie 2012. *Media Perspektiven*, (7-8), 362-379.

Weare, C./Lin, W.-Y. (2000). Content Analysis of the World Wide Web: Opportunities and Challenges. *Social Science Computer Review*, 18(3), 272-292. doi: 10.1177/089443930001800304

Welker, M., & Wünsch, C. (2010). *Die Online-Inhaltsanalyse. Forschungsobjekt Internet*. Köln: Herbert von Halem.

Wirth, W. (2001). Zum Stellenwert der Inhaltsanalyse in der kommunikations- und medienwissenschaftlichen Methodenausbildung. In W. Wirth & E. Lauf (Hrsg.), *Inhaltsanalyse. Perspektiven, Probleme, Potentiale* (S. 353-361). Köln: Herbert von Halem.

Zeller, F., & Wolling, J. (2010). Struktur- und Qualitätsanalyse publizistischer Onlineangebote: Überlegungen zur Konzeption der Online-Inhaltsanalyse. *Media Perspektiven*, (3), 143-153.

Acknowledgments

This work is partially supported by the Collaborative Project FI-CONTENT 2 (www.mediafi.org) funded by the European Commission through the 7th Framework Programme (FP7-603662, Future Internet Public Private Partnership FI-PPP).