Received: Added at production | Revised: Added at production | Accepted: Added at production
DOI: xxx/xxxx

ARTICLE TYPE

A new system for automatic analysis and quality adjustment in audiovisual subtitled-based contents by means of genetic algorithms

Monica Souto-Rico[1,2] | Israel González-Carrasco*[1] | José-Luis López-Cuadrado[1] | Belén Ruíz-Mezcua[1]

[1]Computer Science Department, Universidad Carlos III de Madrid, Madrid, Spain
[2]CESyA, Spanish Centre for Subtitling and Audio Description, Madrid, Spain

Correspondence
Israel González-Carrasco, Avd. Universidad, 20, 28915, Leganés, Madrid, Spain. Email: igcarras@inf.uc3m.es

Abstract

In Spain, the subtitling service on television for the deaf has been improving in quantity since the General Law on Audiovisual Communication was enacted in 2010. This law establishes a series of quality standards that must be followed in the subtitling process. One of the most relevant aspects of subtitle quality is the speed at which they are shown on the screen, due to the fact that a too high speed (less time on screen) will make them difficult to read and the information hard to understand. In order to determine whether the speed at which the subtitles are being shown is adequate, firstly, it is necessary to process all the information associated with the broadcast of the digital TV channels including data from different sources. In this research, the authors have worked with the data obtained within the time period between July 2012 and December 2017, that is, with more than 950 million records. This article presents the ATAD-SUB framework for integration and processing of heterogeneous information associated with the subtitling of audio-visual content from different sources. Moreover, the framework will provide an automatic adjustment of subtitles in broadcasting regarding quality indicators by means of a genetic algorithm approach. The results show that the system is able to estimate the best relationship between the time and size of the subtitles and maintaining the quality levels established for this research. These results have been validated by experts and users of this domain.

KEYWORDS:
Automatic subtitle adjustment, Big Data, Massive knowledge integration, Genetic algorithms

1 | INTRODUCTION

In Spain, the subtitling service on television for the deaf has been improving in quantity since the General Law on Audiovisual Communication was enacted in 2010 (BOE 2010). This law establishes the minimum percentages that accessibility services must fulfill on television in Spain, although it also refers to the technical standards that must be followed for quality subtitling. One of the most relevant aspects in this sense is the speed at which subtitles are shown. Speed means the time in which a subtitle is shown on the screen with respect to the number of characters of the subtitle. In Spain it is measured in "characters per second" (cps) but another metric used in English speaking countries is "word per minute" (wpm).

A high speed (more characters per second) will make it difficult to read the subtitles. It is known that subtitles' speed is partly related to the type of program being broadcast: for example in sports programs, speakers speak very fast in order to report what is happening in the event (Pražák, Psutka, Psutka, & Loose 2013). In order to determine whether the speed at which the subtitles are being broadcast is adequate, it is necessary to process all the information associated with the broadcast of the subtitles. This information is sent through the signal of the television in a continuous flow of signal, and must be processed and extracted to be able to measure it, generating an enormous volume of information. In addition, it is necessary to contextualize this information in order to associate it with the programming that is being broadcast at the time of capturing that information. This information is available through the Electronic Program Guide (EPG) associated with the different Digital Terrestrial Television (DTT) channels. However, the information currently contained in the EPG is not always synchronized with the television broadcast and the information displayed is not accurate , so that in many cases what is described within the EPG are only containers of programs.

Along these lines, there are some quality parameters or indicators (also known as Key Performance Indicator or KPI) that must be taken into account for showing subtitles that are readable and understandable. Within the available set of KPIs, there are three that are of importance for this purpose and which sometimes are in conflict: synchronization, speed and literal subtitling (fidelity or accuracy). Synchronization implies that the subtitle is shown at the same time the speaker is saying what the subtitle contains. As has been stated previously, speed means the time in which a subtitle is shown on the screen with respect to the number of characters of the subtitle and is measured in characters per second (cps). Fidelity means that the subtitle contains exactly the same words the speaker is saying (also known as accuracy). If two of them are prioritized, the third cannot be attended to in those subtitles: for example if synchronization and speed is required, fidelity cannot be assured. This behavior is similar to that presented by distributed systems, where the Brewer (2000) theorem states that a distributed system can only guarantee a maximum of two of the three characteristics presented by these systems.

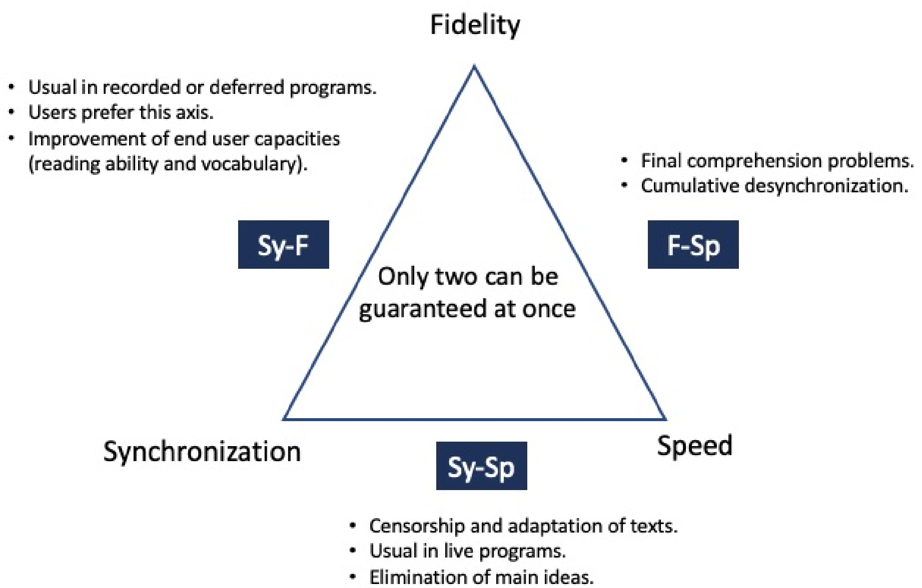


FIGURE 1 Relationships between synchronization, speed and fidelity KPIs (adapted from Brewer (2000))

In this context, as shown in Figure 1, depending on the characteristics to be guaranteed, the perception and understanding of the viewer may vary. Thus, there are three quality indicators that conflict when attempting to include in the subtitles: speed, literal subtitle (fidelity) and synchronization. When two of them prevail, the third cannot be achieved. For example, if for a certain subtitle, quality parameters (KPI) about speed and fidelity are sought (F-SP axe), the synchronization is negatively affected as it inevitably accumulates seconds of delay. Furthermore, and for example, if subtitles of 15cps (speed) with good synchronization (Sy-SP axe) are needed, these subtitles cannot be literal due to the fact that to be within these parameters, it is necessary to economize or modify the locution. Otherwise, if it is intended that the subtitles be literal and synchronized (Sy-F axe), then the speed is an issue since it is related to the speed and pace of the character's speech. Hence, it is important to consider those.

The subtitles in audiovisual content broadcast by television are a challenge from the point of view of quality, especially in live broadcasting. On the one hand, the delay in the availability of subtitles is very disturbing for the consumer of audiovisual content. In this sense, synchronization in live broadcasting is a problem dealt with by the authors in another research study (González-Carrasco, Puente, Ruiz-Mezcua, & López-Cuadrado 2019). At the same time, fidelity can be measured by the NER model, a tool widely used in the quality measurement of subtitles in live broadcasting, where the accuracy has to be checked (Romero-Fresco & Pérez 2015). Therefore, this present research has focused on optimizing the third remaining axis, the speed, but also taking into account the remaining quality parameters or KPIs identified by the authors.

With reference to speed, different research have been carried out to analyze viewer reaction to different television subtitling speed (Jensema 1998). Considering the results obtained, hearing individuals generally wanted slightly slower subtitling and, surprisingly, most deaf viewers (or rather deaf associations) also demand verbatim, and therefore faster, subtitles (Romero-Fresco 2009). Furthermore, the reading speed of viewers is lower than the (average) speech tempo of the original dialogue (Gottlieb 2012). As an example, in Britain, the average reading speed of adult viewers is estimated at 66% of the average speaking speed, according to the Independent Television Commission (De Linde & Kay 2016). Hence, there is no approximation that fits every individual, because many users prefer fidelity over speed or vice versa (Utray, de Castro, Moreno, & Ruiz-Mezcua 2012). As a result, it is necessary to obtain some sort of approximation for television subtitling speed, taking into consideration information about the TV program (genre) and the subtitle elements defined in the reference guide: genre, color or position (AENOR 2012).

With these issues in mind, this article presents a framework for the automatic adjustment of the speed of subtitling in DTT broadcasting, which integrates and processes heterogeneous information associated with the subtitling of audio-visual content. To this end, in addition to the subtitles corresponding to the broadcasts of the different television channels, the information of the programs captured from the EPG is integrated, complementing this information with that available from external sources for programming, generating an enriched EPG in which inconsistencies are detected automatically (e.g., absent information or non-matching information), structuring and breaking down the program information in such a way that the subtitles collected have context information. This information is stored in a big data repository for integration of heterogeneous data from different sources and to allow for further processing. From this repository, the system calculates the speed at which the subtitles have been shown, determining the quality parameters, and associating them to information related to the type of program, which allows further analysis about, for example, the genre of the program (news, sports, entertainment, etc.). Finally, subtitles, speed data and information about the quality parameters or KPIs are sent to the genetic algorithm component in order to determine the best relationship between them.

In this research the authors have worked with the data obtained within the time period between July 2012 and December 2017. The proposed solution is based on the data extracted from the DTT signal as well as the information available in the EPG via web. The information related to the current program as well as the information related to this program is included in the Transport Stream (TS) of the DTT signal (AENOR 2011)(ETSI 2016). Finally, there were more than 950 million records available for this research that are stored in the Big Data Persistence Layer of the framework. The first step of the ATAD-SUB framework is the gathering and preprocessing of the information recovered from the DDT broadcasting and the enriched EPG. The idea is to integrate into the framework the heterogeneous information associated with the subtitling of audio-visual content. Moreover, the framework proposed incorporates a genetic algorithm based on NSGA-II in order to optimize the time and size of the subtitles maintaining the quality defined by the KPIs. NSGA-II is a genetic algorithm for multi-objective optimization with low computational requirements, an elitist approach, and parameter-less sharing approach (Deb, Agrawal, Pratap, & Meyarivan 2000). As has been stated, in the context of the present research, there are conflicts between two objectives, speed and quality, which will make improvement in one of them lead to a worsening in the other. In this study, different scenarios have been included to test the proposal: deferred, semi-live (with live and deferred parts) and live programs. In this evaluation process, the optimal time and size for subtitles has been determined considering the quality parameters of the KPIs. Next, the results obtained have been validated by final users with hearing impairments by means of questionnaires based on the Likert scale (Boone & Boone 2012). In all of the scenarios, the framework is able to gather and process heterogeneous information and return as output the optimal speed regarding quality parameters by means of the multi-objective genetic algorithm NSGA-II.

In summary, our main goal is to integrate heterogeneous information in a big data domain to provide an automatic adjustment of subtitles in broadcasting regarding quality indicators by means of a genetic algorithm approach. Therefore, the expected contribution of this research is a solution that improves the quality of life of people with disabilities when consuming audiovisual contents. Additionally, there are other users who can directly benefit from this system: children, the elderly, persons with intellectual disabilities, people learning languages, since the possibility of improving reading and writing skills is an added value to subtitling services. In addition, subtitling solutions like that presented here, can also be of benefit to the general public in certain noisy environments (public transport, commercial areas, etc.), and their knowledge of them could help sensitize and prepare them for possible age-associated hearing loss. Furthermore, using this technology for distribution in television or streaming on the Internet can be a key contributor to the education sector (González-Carrasco et al. 2019).

The rest of the paper has been structured as follows. Section 2 outlines the relevant literature in the area of genetic algorithms and the issues associated with the use of accessibility services based on subtitles. Section 3 describes the main issues of the problem to be solved. Section 4 discusses the central features of the framework proposed, including a usage scenario and the main components of its architecture. Section 5 describes assessment of this proposal. This section also presents the main results, the methodology used for expert validation and a final discussion. Finally, in Section 6 the paper ends with a discussion on research findings, limitations and concluding remarks.

2 | LITERATURE REVIEW

Synchronization, speed and fidelity of subtitles has been a subject of study for previous researchers, who have identified the main features and issues in this area. Specifically, the automatic synchronization of subtitles and audio has been examined by a significant number of studies in this area. These synchronization proposals are commonly considered a component-tool for the process of automatic subtitle generation (González-Carrasco et al. 2019). For programs with simple environments such as news programs, the literature offers a wide range of techniques for the automation of synchronization (Gao, Zhao, Li, & Yan 2009)(Ando et al. 2003)(Garcia et al. 2009)(Lertwongkhanakool, Punyabukkana, & Suchato 2013).

The speed of subtitles and its relationship with their comprehension has been under study during recent years. For example, Romero-Fresco (2010) has focused research on comprehension and viewing patterns of subtitled news, trying to assess whether it is related to the speed of the subtitles or to other factors. Along these lines, there are many other people within the target audience for subtitles (for example, the elderly), who have limited reading speed and for whom subtitles with more than 180 words per minute produce frustration (Romero-Fresco 2009). Moreover, different tools has been developed in order to make reading and estimations of speed calculations, and whose results are expressed by means of CPS (characters per second) and WPM (words per minute) (Ferriol 2013a)(Ferriol 2013b). Another study set out to evaluate deaf people's experiences of subtitling, looking specifically at how comprehension and enjoyment of different types of programs are affected by the speed of the words on screen. For people who utilize more than one sensory input, subtitles are often used as a useful reference point if they miss a word, so speed is not such a critical factor. For those who rely more heavily on subtitles, speed is much more of an issue because if they cannot read them fast enough they will literally lose the plot (Ofcom 2005). Viewer reaction to different television captioning speeds has been analyzed by Jensema (1998). In this study, participants adapted well to increasing caption speeds. Most apparently had little trouble with the captions until the rate was increased to at least 170 WPM. Hearing individuals wanted slightly slower captions. However, this apparently related to how often they watched captioned television. Frequent viewers were comfortable with faster captions. Age and sex were not related to caption speed preference; nor was education, with the exception that people with a graduate school level of education showed evidence that they might prefer slightly faster captions. In the case of live subtitling, and more specifically respeaking, the most common method used to evaluate the quality of subtitles produced in real time consists of assessing their accuracy. In this context, fidelity or accuracy of the subtitles can be measured by the NER model, a tool widely used in the quality measure of subtitles in live broadcasting, where the accuracy has to be checked (Romero-Fresco & Pérez 2015).

Regarding the transmission of the audiovisual contents, in DTT broadcast and IPTV, TV channels are delivered to users according to MPEG-2 and DVB Transport Stream codification techniques. Subtitles can be conveyed in the Transport Stream in the form of DVB Subtitle stream(s) and/or as Teletext subtitles embedded in the Teletext stream. Everything is assembled according to MPEG-2 and DVB standards to make up a multimedia service (the MPEG-2 term for TV channel) that is transmitted over IPTV or DTT broadcast networks. The use of a common time base is essential to the process of putting together video, audio and data (subtitles are a special type of data) in the same Transport Stream and every transmitted packet specifies its presentation timestamp, regardless of its type(video, audio or data) (de Castro, Carrero, Puente, & Ruiz 2011) (ETSI 2016).

The ATAD-SUB framework allows real time storage for DTT subtitles. In this sense, Opensubtitles is one of the best-known services that contains one of the largest free online subtitle database. This platform supports several languages and also provides an API for developers. The Opensubtitles platform has been used for different research: for defining a statistical machine translation of subtitles (Müller & Volk 2013), for defining a corpus of Interaction-Response pairs extracted from subtitles files, created to help dialogue systems to deal with Out-of-Domain interactions (Ameixa, Coheur, & Redol 2013) and for detecting Cross-Lingual Semantic Divergence for Neural Machine Translation (Carpuat, Vyas, & Niu 2017). However, this platform is focused on TV series and movies, and no subtitle storage services have been found associated with TV broadcasts and real time processing.

Taking into account the amount of information related to the research in this area, with the emergence of Big Data, the use of NoSQL (Not only SQL) technology is rising rapidly among internet companies and other enterprises. NoSQL databases are increasingly considered a viable alternative to relational databases, as more organizations recognize that its schema-less data model is a better method for handling the large volumes of structured, semi structured and unstructured data being captured and processed today (Bhogal & Choksi 2015). The NoSQL landscape presents several possibilities and implementations, highlighting among them MongoDB (Corbellini, Mateos, Zunino, Godoy, & Schiaffino 2017). MongoDB is a Document-oriented database, widespread in the industry with support for queries on multiple fields (Chodorow 2013).

In recent times, genetic algorithms have been used for searching among alternatives, maximization / minimization problems like the traveling salesman problem, sorting problems, multi-objective decision making, multi-criteria decision making and constrained optimization problems (Kar 2016)(Grefenstette 2013).

In the context of large volumes of data, bioinspired algorithms may enable hard real-world problems and applications to be successfully faced in the Big Data context (including but not limited to social data analysis) (Camacho et al. 2018). Genetic algorithms are increasingly being used for large scale problems like non-linear optimization (act of achieving the best possible result under given circumstances) clustering and job scheduling Verma, Llorà, Goldberg, and Campbell (2009). Moreover, genetic algorithms have been used for different purposes related to audiovisual contents and text segmentation. Lamprier, Amghar, Levrat, and Saubion (2007) develop a new algorithm for linear text segmentation on general corpuses, and Uğuz (2011) includes genetic algorithms for text categorization problems. Moreover, Song and Park (2009) and Song, Li, and Park (2009) define different solutions for text clustering using genetic algorithms. Huang, Hsu, and Sandnes (2007) define an Intelligent Subtitle Detection Model for Locating Television Commercials. Kundu, Das, and Bandyopadhyay (2013) describe a mechanism for fragmenting a sequence of movie script dialogue into scene-wise groups. In other words, it attempts to locate scene transitions using information acquired from a sequence of dialogue units.

Additionally, multiobjective genetic algorithms have been defined for solving several problems in different domains. Problems comprising several conflicting objectives have sprung a vibrant research activity during the last 20 years, with an ever-growing body of literature, competitions, and benchmark focused on deriving new bio-inspired solvers suited to produce Pareto optimal solutions with increased efficiency and efficacy (Del Ser et al. 2019). In the context of text processing, the NSGA-II algorithm has been used for text summarizing (Sanchez-Gomez, Vega-Rodríguez, & Pérez 2018) (Fors-Isalguez, Hermosillo-Valadez, & Montes-y Gómez 2018) or for clustering of medical text (Sahoo, Ekbal, Saha, Mollá, & Nandan 2016).

3 | BACKGROUND AND PROBLEM DESCRIPTION

The term audiovisual content broadly refers to “any dual media presentation consisting of visual, usually moving, images or pictures, together with auditory, and sometimes printed, language". Audiovisual input involves the overall interplay of two channels, the visual and the auditory one and the interaction of five different tracks, i.e. moving image, voices, sounds, music, and written text on the screen (Ghia 2012)(Baggett 1989). Furthermore, subtitling is a special form of audiovisual translation, owing to characteristic features and a number of genre–specific constraints which come into play. These involve spatial, temporal, synchronization and technical–perceptual issues linked to subtitle distribution and readability (Gambier 2003). In this section, the authors present relevant topics in the subtitling process for the broadcast of audiovisual content, the current Spanish legislation, and the main quality parameters defined in this domain.

3.1 | Television subtitle services

Television subtitle services can be classified according to the following categories (González-Carrasco et al. 2019):

- When are they generated (before or during the broadcasting)?
- How are they generated (typing, stenography, re-speaking)?
- When are they introduced (in a predefined way, live synchronized, as soon as possible)?
- How are they broadcasted (incrusted, linked, synchronized or unsynchronized)?
- What type of program are they used in (pre-recorded, scripted, live improvised, mixed)?

In this research, the subtitles obtained during the broadcasting of the audiovisual contents can be placed in any of these categories, so several combination of subtitling scenarios can be found.

3.2 | Spanish Audiovisual Legislation

The Spanish General Law on Audiovisual Communication was published in 2010 BOE (2010). This law indicates the percentages of TV programming that accessibility services must have for television in Spain: 90 percent for public channels and 75 percent for private channels (all of them with national coverage). Although this Law explicitly states that the accessibility services broadcast on television must follow accessibility guidelines, according to the technical regulations in force at that time, how this quality is accounted for has been left open.

At present, the only technical regulation where the quality of the subtitles is examined and defined is the UNE 153010 published in 2012 by AENOR (AENOR 2012). This guideline was published in 2012 and it establishes different recommendations and rules for subtitling for the deaf and hearing-impaired in the Spanish audiovisual domain. It is important to note that the Audiovisual Communication Law indicates that this UNE regulation should be followed, but not the percentage of quality that the subtitles should have, that is, if they must meet all the requirements, or only some or the minimum percentages that a service must have to consider it a quality service. In this sense, and in order to define different quality requirements for a subtitle, the authors have considered some basics characteristics of the UNE 153010 guideline for defining quality parameters or performance key indicators.

3.3 | Key Performance Indicators

A Key Performance Indicator or KPI is a measurable value that demonstrates how effectively a key goal is achieved. Therefore, KPIs are used to help in measuring the progress toward achieving strategic goals in many domains Parmenter (2015). Taking this into account, the authors have defined some quality parameters (KPIs) for the subtitling process performed by audiovisual broadcasting in Spain considering technical regulations, professional experience and industry recommendations and guidelines.

- **Visual aspects: position, font size, line length and typography**. Subtitles for the deaf have many characteristics related to visual aspects like the position, font size and font type and the length of the subtitle. The typography features are usually defined by the receiver or the broadcaster in the case of television in Spain. The length should not exceed 37 characters per line, including spaces and punctuation marks. Font size should be set for allowing the emission of subtitles without exceeding the screen security standards (screen pixels). Finally, for visual aspects, the following criteria have been established:
  - Deferred subtitles cannot hide information.
  - Lip sync is fundamental (for synchronization issues and for users who employ lip reading).
- **Temporal aspects: subtitle speed and synchronization.** The main indicators of quality in terms of temporal aspects are the subtitles' speed and synchronization. The subtitles should be perfectly synchronized with all the audio information of the content (dialogue, sound effects or songs). This indicator only applies for deferred contents, since currently the technology presents several issues for the synchronization in live events (González-Carrasco et al. 2019). Regarding speed, one quality indicator is that any subtitle, regardless of its length, must have a duration of at least 1 second. Nowadays, language-saving techniques are used to achieve the 15cps stipulated as quality speed. This faces with requests by users for a literal subtitle.
- **Characters identification: Color and labels.** One of the most important indicators when measuring quality in a subtitle for deaf people is identification of the characters. This identification can be done by using colors, no more than five, and labels. Using the same colors throughout the whole content to identify the characters is recommended.
- **Sound effects and contextual information.** These indicators are only feasible in deferred contents. The sound effects are important for the good tracking of the plot. These are presented in the upper right part of the screen in parentheses and with the first letter in uppercase. The contextual information indicates the way the character is expressing her or himself, giving information about the tone of the phrase. This information must be presented before the sentence it affects, in parentheses and in capital letters. Voice-over is also treated as contextual information, at the beginning of the phrase they accompany, in parentheses and in uppercase.
- **Music: format and differentiation between music and songs.** These are valuable indicators in deferred subtitles. While the music is presented as a sound effect, the songs (with their lyrics) belong to the group of subtitles located at the bottom of the screen. The music subtitle must indicate at least one emotional aspect, while the songs will only be subtitled if they are important for the plot (using the color of the corresponding character). In addition, lyrics must be accompanied by a # in each subtitle line.

- **Editorial criteria: grammar, spelling and fidelity (accuracy).** The last quality indicator measures the absence of spelling errors, except those inaccuracies that are a character's leave or a contribution to the work and that will be subtitled in quotes. According to the users, the subtitles must always be literal with the speech (Romero-Fresco & Pérez 2015), except for live ones because of a technological issue.

Within the set of KPIs defined, there are three of them that are important for this purpose and sometimes are in conflict: synchronization, speed and literal subtitle (fidelity or accuracy). As has been stated in the introduction, this research has focused on optimizing the speed, but also taking into account one quality parameter or KPI identified by the authors in this section: line length. Therefore, the categories involved are temporal aspects (speed subtitles), but also visual aspects (line length) for quality parameters.

3.4 | Problem description

The subtitles in audiovisual content broadcast by DTT are a challenge from the point of view of quality, especially in live broadcasting. Synchronization, synchrony between the subtitles and the audiovisual contents, is basic for providing a quality subtitling service as a support for content understanding. Also, synchronization helps understanding subtitle text regarding the image, especially in news programs (or similar), where audiovisual content is shown at such speed that if synchronization is not maintained, the subtitles may appear on screen with other completely different visual information, generating confusion among users. At the Spanish level, the "Follow-up report of the subtitles and audio description in TV" (Ruiz et al. 2014) states that one of the main complaints made by TV' s users is the lack of synchronism between audio and subtitles in some types of programs (González-Carrasco et al. 2019). Moreover, delay remains one of the most significant factors in the audience's perception of lack of quality in live-originated TV captions or subtitles for the deaf and hearing-impaired (Simpson, Barrett, Bell, & Renals 2016).

According to the technical regulation UNE 153010 (AENOR 2012), by fidelity or accuracy it is understood that the subtitles are in accordance with the phrase and they have a proper sense. In no case does literal subtitling interpret the figurative meaning of the phrase. For the deaf user, high fidelity is important to ensure that the ideas that are heard are not edited and as such, there is no censorship in the information received. In 2016, the Spanish Centre for Subtitling and Audio Description conducted a study on the preferences of users with disabilities regarding the speed and fidelity of the subtitles. It came to the conclusion that comprehension decreased when the subtitles were summarized to achieve a lower speed of subtitle exposure, that is, a subtitle close to the 15cps set by current regulations. In fact, the subtitle user is able to accommodate a wide range of speeds that have improved over the years (Ofcom 2005). Therefore, it can be concluded that in the production of subtitles for audiovisual contents fidelity over speed must prevail (AMADIS 2016). Subtitle fidelity regarding audiovisual contents can be measured by the NER model, a tool widely used in the quality measurement of subtitles in live broadcasting, where the accuracy has to be checked (Romero-Fresco & Pérez 2015). In deferred contents, the subtitles should have total accuracy because the production of the subtitles is not a real time process. Accordingly, in this case, the subtitles should be totally literal for a quality subtitling service.

The most important aspect regarding these three indicators is that they cannot occur at the same time. The "subtitler" (the person in charge of the subtitling process) must prioritize two of them, so the third one will always be compromised or will not be taken into account. If literal subtitles are intended and a speed of 15cps is set, synchronization is complex and the process will accumulate delays of seconds in order to prioritize fidelity. On the other hand, when prioritizing speed and synchronization, the generated subtitles cannot be literal or are not able to specify everything that is transmitted aurally. Subtitling is a service that should serve both the totally deaf user and one who has hearing loss and only uses the subtitles as a support tool. For this latter type of user, fidelity is more important than the other features, since they have some hearing ability. If the subtitles are not literal, users may understand that there is a certain type of censorship or some valuable information is missing that is necessary for total understanding. Furthermore, fidelity and synchronization can be prioritized to avoid losing information, but in this case the speed in very few cases would be 15cps as specified by the standard. In some genres labeled as documentaries, a much lower speed can be found, but not in most television genres where speed prevails when giving information. Users' preferences in Spain opts for these two indicators, fidelity and synchronization, leaving aside the speed (AMADIS 2016), since the users' reading skill improves with time and with the (previous) knowledge about the audiovisual content.

Finally, one of the points made by the Spanish technical regulations is that the subtitling service is not only intended for users who are deaf or who have hearing disabilities, but also benefits users who are learning to read, foreigners who want to improve language skills or even those users with intellectual disabilities who need support to better understand the contents. In addition, the population is getting older with the hearing difficulties that this may entail. Moreover, the reading of subtitles improves skills of the users, for example in children, and the access to information, so that at a higher speed of reading and understanding of the text, improved understanding of the audiovisual content can be achieved.

4 | SOLUTION PROPOSED

4.1 | Data description

The proposed solution is based on the data extracted from the DTT signal as well as the information available in the EPG via web. In DTT, audiovisual contents and additional information is delivered by means of the signal sent from the broadcaster to the receivers. Thus the DTT signal must be processed in order to separate on the one hand the audiovisual content, and on the other hand the additional data related to the emission. The relevant data for this research are EPG information and the accessibility elements (subtitles in this research). The DTT signal also contains information regarding the time in which each element should be shown on the screen.

The content of the DTT signal is complex and must be extracted and processed in real time in order to obtain the information about the current program to be shown. It is remarkable that the process is in real time, due to the fact that the information is transmitted and should be processed at the moment. Usually this task is performed by DTT receivers, who process the DTT signal, extract the audiovisual contents and show the synchronized audiovisual content as well as the rest of the elements.

The information related to the current program as well as the information related the accessibility elements is included in the Transport Stream (TS) of the DTT signal. The content of the TS is described in norm ETSI EN 300 468 (ETSI 2016). This norm is translated to the Spanish environment by norm UNE 133300 (AENOR 2011).

Analyzing the content of the TS during the time in which a program is broadcast, the subtitles can be retrieved, as well as their characteristics and the time they were shown. Based on these times the speed at which subtitles are shown can be determined.

However, it is well-known that the speed of the subtitles is highly related to the type of program. For this reason and for the purposes of our research, it is also necessary to determine the name of the program, the time at which the program was broadcast and the content descriptor of the program. The content descriptor of the program is a word or set of words that describes the content of the program. For example, "sports" is the descriptor used for broadcasting basketball games. In the context of this study, the term "genre" is used for referring to this concept.

The information of the program is usually included in the TS. For this reason, the relevant data for this research is based on the following elements retrieved from the TS.

- Program name. Name of the program.
- Start Time. Time at which the program starts
- End Time. Time at which the program ends.
- Duration. Duration of the program.
- Genre. Content descriptor of the program. The values for this element are defined by ETSI EN 300 468 ETSI (2016) .
- Subtitles. For each subtitled program, each subtitle line is stored. A subtitled line is a text line shown in the screen. Sometimes you can see two lines of subtitles in the screen. Each line is stored in separate records in the database. The information stored about each line is:
  - Text. Text of the line.
  - Position. Position of the text line in the screen.
  - Color. Color of the text. The colors of the text can be 'White', 'Yellow', 'Magenta', 'Green', 'Blue', 'Cyan', 'Red', 'Black'
  - Presentation Time Stamp (PTS). Time at which the subtitle is shown on the screen (in ms.).
  - Hiding time. Time at which the subtitle is removed from the screen (in ms.).

Norm ETSI EN 300 468 (ETSI 2016) defines the content descriptors for each program, so the possible values of the field genre are: Movie/Drama, News/Current affairs, Show/Game show, Sports, Children/Youth programms, Music/Ballet/Dance, Arts/Culture (without music), Social/Political issues/Economics, Education/Science/Factual topics, Leisure hobbies.

All this information is available in the TS. However, sometimes the information related to the programs is not precise (and sometimes it is missing). For this reason, a web service is required for getting additional information from the broadcasters. Usually broadcasters include information about the programs and their schedule in the web, so that it can be used for fixing incomplete information of the EPG. From the web services the name of the program, and start and finish time is obtained for each program: this information is sufficient to complete the information of the EPG.

Once the information has been extracted from the signal, it is processed in order to obtain derived quality information. Once the information of programs and subtitles has been merged, the data related to each subtitle can be represented as a tuple *S*

$$S_i = (Text_i, L_i, Pos_i, Col_i, Tshown_i)$$

Where:

- Text_i: is the text of the subtitle i.
- L_i: is the length of the subtitle i. This length is calculated based on the text extracted from the TS for each subtitle.
- Pos_i: is the position of the subtitle in the screen i.
- Col_i: is the color of the subtitle i.
- Tshown_i is the number of seconds that the subtitle lasted in the screen. This time is calculated as the difference between the timestamp when the subtitle was shown and the timestamp when the subtitle was hidden.

Since each subtitle is in the context of a program, the information related to the genre is available.

4.2 | Framework architecture and steps

According to the state of the art review and included in this paper, the subtitles in audiovisual content broadcast by television are a challenge from the point of view of quality, especially in live broadcasting. Three main features have been defined for quality subtitles: synchronization, fidelity and speed.

Synchronization and fidelity have been dealt with by other research and different approaches have been analyzed for different scenarios: live and deferred subtitles. Moreover, today one interesting issue to be resolved is obtaining some sort of approach for television subtitling speed taking into account information about the TV program (genre) and the subtitle elements defined in the guidelines and regulations.

This research focuses on the adjustment of time and size of the subtitles considering quality features as a key element to guarantee a high-quality accessibility service for people with hearing disabilities. To do this, the ATAD-Sub framework enables automatic extraction and analysis of audio-visual content and subtitles from the DTT signal, using a big data repository for integration of heterogeneous data from different sources.

The proposed framework is structured into a set of components in charge of the process of acquiring the information from external sources, storing and processing it to obtain the most adequate quality parameters for subtitled-based contents. Figure 2 shows the different components and steps of the process developed by the ATAD-SUB framework for automatic analysis and quality adjustment in audiovisual subtitled-based contents by means of genetic algorithms. The first step of the ATAD-SUB framework is the gathering and preprocessing of the information recovered from the DDT broadcasting and the enriched EPG (gathering and pre-processing components). The idea is to integrate into the framework the heterogeneous information associated with the subtitling of audiovisual content (data management component). Moreover, the framework proposed incorporates a genetic algorithm based on NSGA-II in order to optimize the time and size of the subtitles maintaining the quality defined by the KPIs (optimization component).

The steps of the framework can be summarized in:

1. Data extraction from DTT signal. The Gathering component extracts the subtitles and information about the TV programs from the DTT signal.
2. The extracted data is preprocessed by the preprocessing component in order to complete the information related to the TV programs using web resources and associate the subtitles with its corresponding programs.

3. The data is stored in a big data repository, including calculated quality information in order to exploit it. The Data management component performs these operations and makes this data available to the optimization component.

4. The optimization component is in charge of analyzing the content of the subtitles and proposing improvements. This component reads the subtitles from the data management component and proposes alternative subtitles according to the KPIs related to the time on screen and the size of each subtitle.

Next subsections describe in deeper detail each of the components.

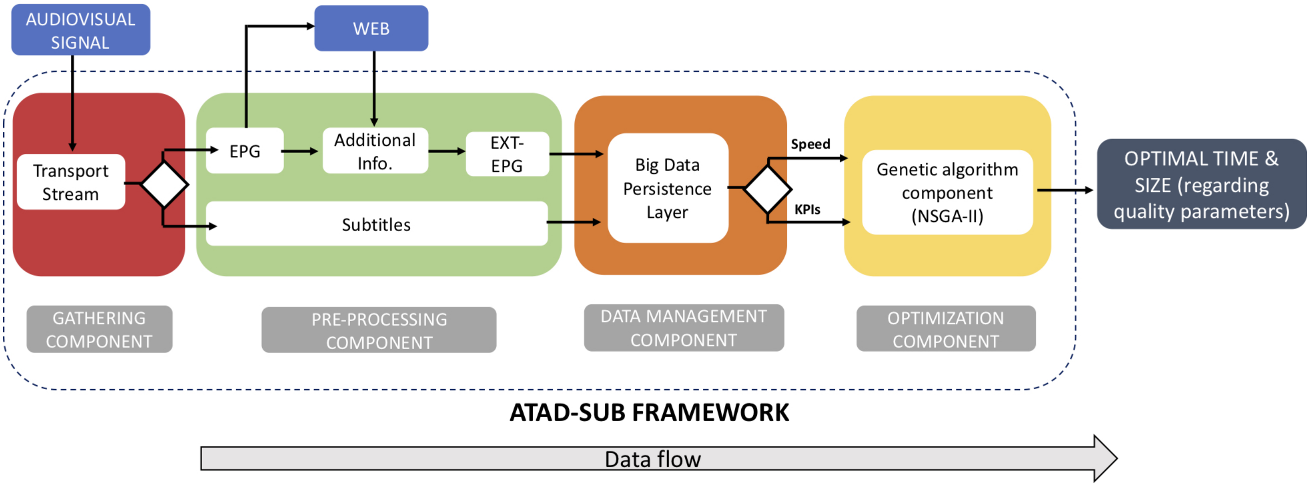


FIGURE 2 Components and relationships in ATAD-SUB framework

4.2.1 | Gathering component

DTT broadcasters send the audiovisual content encoded into a signal by means of antennas. This signal contains a TS with the audiovisual content as well as additional information. As described in section 4.1, along with the audiovisual content broadcasters include the EPG information and the subtitles of the program. The complete content of the signal includes a wide range of elements; however for the purposes of this research only subtitles and EPG information are relevant. Since the signal is broadcast as a continuous flow, the information should be extracted and stored in order to be processed.

The gathering component is connected to a DTT antenna in order to receive the signal broadcast by channels. This component is continuously monitoring the content of the signal of national channels, analyzing the TS, retrieving the EPG information and storing the subtitles. Both EPG information and subtitles are stored into separate text files. It should be highlighted that the information of the program and the information related to the subtitles are not together in the signal. For this reason, they are firstly stored separately in order to streamline the signal processing.

4.2.2 | Pre-processing component

The information of the EPG extracted from the DTT signal and the broadcast subtitles are available, but they are in separate records. Then it is necessary to relate each subtitle to the information of its corresponding program.

The first problem to be addressed is that EPG information is sometimes missing or incomplete in the TS. For this reason, the gathering component retrieves the information of the EPG provided by the channels from the web services. For each program extracted from the signal, this component retrieves the available information of the channel from the web service . Then, the component compares the EPG received in the signal with the information retrieved from the internet, and unifies the information in order to determine for each channel what program has been broadcast in a given moment as well as its genre. If the program information was missing in the signal, the component uses the information retrieved from the internet. If the information of the web is more detailed than the information of the signal, the component updates the signal information with the one retrieved from the internet: for example in the EPG a TV series is programmed from 10 am to 11 pm, but in the internet the information could be more precise including information about two episodes, one from 10:00 am to 10:30 am and another from 10:30 am to 11:00 am. The result of merging this information is the extended EPG.

The second problem is relating each subtitle with its corresponding program. First of all, time information of subtitles and EPG is translated into Unix Time Stamp format (which is the number of seconds since Jan, 1st 1970) in order to facilitate the time calculations. In this way, the time of each subtitle can be calculated in seconds and these values can be used to calculate the subtitles speed.

Since for each subtitle the presentation time and the channel are stored, it is possible to determine the program, because only one program can be shown in a channel at a given presentation time. This information is stored by the Data Management Component.

4.2.3 | Data management component

Data extracted from the signal and the web by the extraction component, and processed by the pre-processing component should be stored in a proper way in order to allow further optimization of their quality attributes.

In the previous steps, data is stored in text files in order to optimize the online processing. However, at this point the use of big data techniques is necessary so as to improve the performance and facilitate the process. The use of a big data environment is not related to data complexity but to the huge amount of data to be stored.

The big data repository is available for the optimization component. Since the EPG information and the subtitles are not yet related, the first process on the stored data consists of relating each subtitle to its corresponding program. It is possible because each subtitle has a timestamp recording the moment when it was shown and also is related to a channel. Since in the extended EPG generated by the pre-processing component, only one program can be broadcast by a channel at a given time, programs and subtitles can be related. Furthermore, some additional information must be derived from the subtitles. Regarding the KPIs to be considered in this research, subtitle color, subtitle length and subtitle speed must be calculated.

Since the timestamp of presentation and hiding is stored for each subtitle, and the number of characters of the subtitle can be measured counting each character of the line, the speed of the subtitle is calculated based on the following formula:

$$Speed_i = \frac{Len(S_i)}{H_i - P_i} \tag{1}$$

Where

- Speed_i is the speed of the subtitle i (measured in characters per second)
- $\mathrm{Len}(S_i)$ is the number of characters shown in the subtitle i.
- H_i is the time when the subtitle S_i was hidden.
- P_i is the time when the subtitle S_i was presented.
- $H_i - P_i$ is the number of seconds between H_i and P_i.

Finally, the color is represented as a number based on its RGB value in order to facilitate the calculations of the KPIs. All this information is available for the optimization component.

4.2.4 | Optimization Component

The well-known efficient multi-objective generic algorithm, called NSGA-II is applied in this component (Deb, Pratap, Agarwal, & Meyarivan 2002). The purpose of the Optimization Component is to identify the best set of subtitles according to the quality KPIs established for this research. In this way, the obtained subtitles should achieve the best balance between speed and length in order to improve the readability for people with disabilities, assuming that the fidelity is constant.

The problem to solve is defined as a set of words to be shown with optimal subtitles in length and speed. The set of words is defined by the existing subtitles stored in the data management component. Each word has a length and a time in which it must appear on the screen (assigned to the original subtitle). Since the aim of the optimization component is to obtain optimized subtitles, the first step of the optimization consists on put together all the possible subtitles in order to split them again in a more optimal way according to the proposed KPIs. To do so, the system applies heuristic rules in order to determine the set of words that can appear together. These heuristics are summarized in:

- Words that appear in the same subtitle.
- Words that appear in consecutive subtitles with less than 1 second between the hiding of the first subtitle and the showing of the second. If the color of the subtitles are different they cannot be considered together, because it means that it correspond to different speakers.

As described in section 4.2.3 the optimization component will receive a number of subtitles with their characteristics and KPIs. The goal is to optimize the subtitles according to the KPIs related to the number of the subtitle characters and the speed. To do so, the framework needs to calculate the subtitle lines again. For this reason, the first step for the optimization consists of putting together all the text to be shown in the subtitles. In order to maintain a reasonable size of the chromosomes, and following the rules of the generation of subtitles, the text will be limited to those subtitles that must be shown together. Two subtitles Si and Si+1 must be shown together (which means consecutively) when $H_i - Pi + 1 < 1second$ and $Col_i = Col_{i+1}$ (because if they have different colors, the subtitles cannot be merged since they usually correspond to different characters). Subtitles with more than one second of difference cannot be processed together because it means that the text of the subtitle will not be coordinated with the sound. Finally, subtitles with different colors correspond to different characters and so cannot be merged. As a result, the framework obtains a text and the lapse of time in which this text must be shown: this time is defined by the time P_1 of the first subtitle and the H_n of the last subtitle to be processed.

With the set of subtitles that can be optimized, the system obtains the number of words. For each word the system calculates its length and a time frame corresponding to the subtitles of which they form part. In this way all these words should appear in a time frame defined by the showing time of the first subtitle and the hiding time of the last one. Since the system can calculate the total number of characters and the time frame, the average speed for each character can be calculated in order to be considered (if necessary).

Based on the approach of Kundu et al. (2013) and Lamprier et al. (2007), each chromosome is defined as serial of 0 and 1. The chromosome has as many 0 and 1 as words to include in the subtitles. An 0 represents that after the word the subtitle line can continue, and 1 indicates that that is the last word of the subtitle.

Let's suppose a set of words $W = "La", "casa", "se", "quedo", "vacia"$. The chromosome $C = 00100$ produces the following subtitles:

Line1: "La casa se"
Line2: "quedó vacía"

The last bit is unnecessary since it will always be 1 (because after the last word the subtitle always ends), so the chromosome size is n-1 where n is the number of words to be displayed. Given this chromosome structure, each individual represents a possible subtitle distribution for each group of words that should be displayed together.

In Figure 3 the interaction between the data management component and optimization component of ATAD-SUB framework is depicted. From the data persistence layer, the framework extracts different consecutive subtitles in order to feed the optimization component. The optimization component generates the different chromosomes for optimizing the subtitles regarding time and size.

Figure 4 shows an example (in Spanish) of one Consecutive Block of Subtitles (CBS) obtained from the data management component of the framework. This CBS is created by merging different subtitles obtained from the DTT that have been shown consecutively in the original broadcasting. The total time in seconds of the subtitles is also included in the CBS.

In each population, the framework integrates the chromosome, the different lines of subtitles and the objective functions time (F1) and size (F2). In the chromosome a value of "1" indicates a cut in the CBS subtitle because it is the last word in the line. In Figure 5, an example of two members of the population obtained by the genetic algorithm in the optimization component of ATAD-SUB framework is shown. Despite the fact that the text

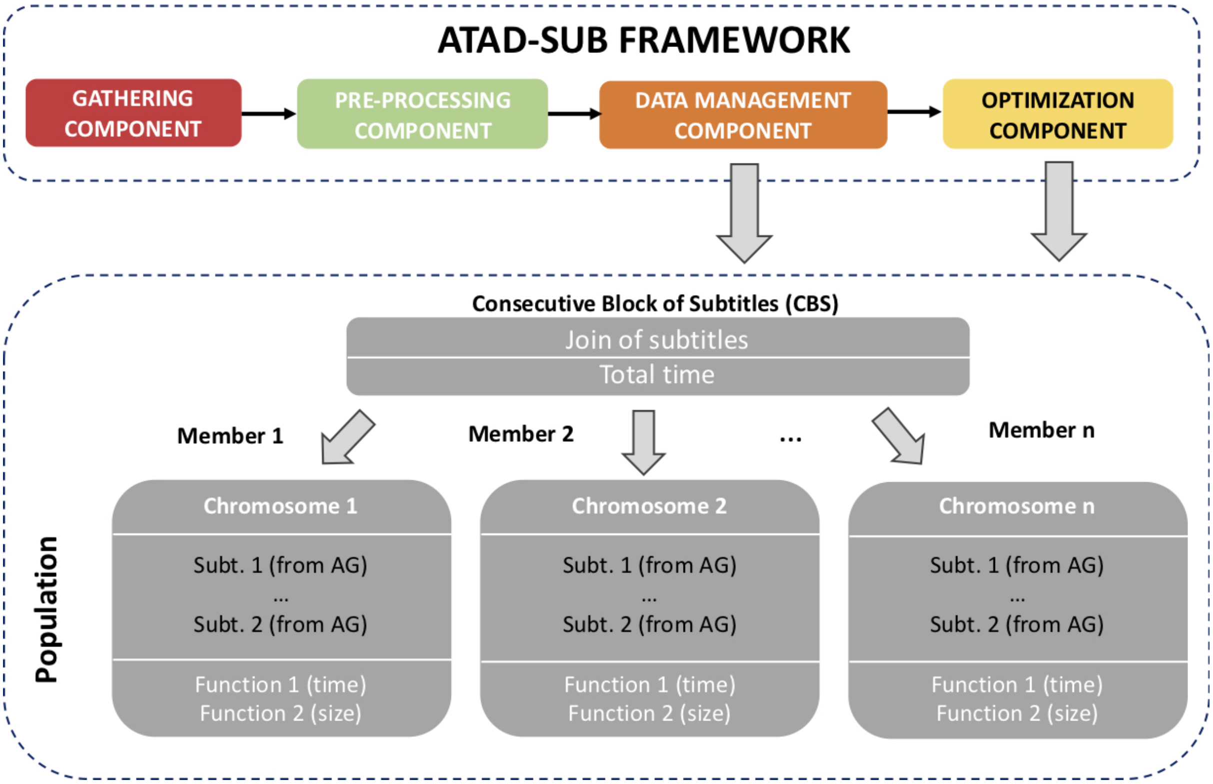
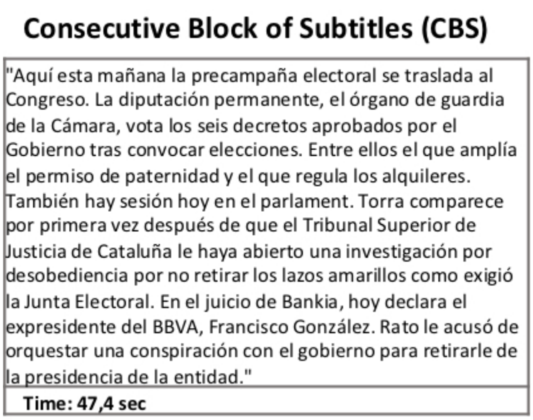


FIGURE 3 Interaction between data management component and optimization component


Consecutive Block of Subtitles (CBS)

"Aquí esta mañana la precampaña electoral se traslada al Congreso. La diputación permanente, el órgano de guardia de la Cámara, vota los seis decretos aprobados por el Gobierno tras convocar elecciones. Entre ellos el que amplía el permiso de paternidad y el que regula los alquileres. También hay sesión hoy en el parlament. Torra comparece por primera vez después de que el Tribunal Superior de Justicia de Cataluña le haya abierto una investigación por desobediencia por no retirar los lazos amarillos como exigió la Junta Electoral. En el juicio de Bankia, hoy declara el expresidente del BBVA, Francisco González. Rato le acusó de orquestar una conspiración con el gobierno para retirarle de la presidencia de la entidad."

Time: 47,4 sec

FIGURE 4 Example of Consecutive Block of Subtitles

is the same for all the chromosomes, it can be shown how the different chromosomes (combinations of 0 and 1) produce a different distribution of subtitle lines, each one with its own size and time. Since each line has a different size, the F1 and F2 values for each chromosome will be different, because they do not depend on the whole text but on the distribution of this text in different lines.

Thus, the framework receives as input a text to be separated into several lines of subtitles and the lapse of time (in seconds) that the text must appear on the screen. With the number of words, the chromosome size is established and a population is created randomly based on the chromosome size. When a chromosome is created, the framework calculates the number of subtitle lines (it is equal to the number of "1" in the chromosome plus one), and assign to each line its corresponding words according to the rules described above. The chromosome also receives the time available for the subtitles. Once the lines has been created for the chromosome, it is necessary to calculate the number of characters for each

Member 1 of population

'00000010000001000010000010000010000000100000010000010100000000100001000010000100000000010000 001000001000001000010000000'
Aqui esta mañana la precampaña electoral se traslada al Congreso, La diputación permanente, el órgano de guardia de la Cámara, vota los seis decretos aprobados por el Gobierno tras convocar elecciones, Entre ellos el que amplía el permiso de paternidad y el que regula los alquileres, También hay sesión hoy en el parlament, Torra comparece por primera vez después de que el Tribunal Superior de Justicia de Cataluña le haya abierto una investigación por desobediencia por no retirar los lazos amarillos como exigió la Junta Electoral, En el juicio de Bankia, hoy declara el expresidente del BBVA, Francisco González, Rato le acusó de orquestar una conspiración con el gobierno para retirarle de la presidencia de la entidad,
F1: 1,337
F2: 0,103

Member 2 of population

'0000001000000100001000001000001000000010000001000000010000000010000100001000010000000001000 0001000001000001000010000000'
Aqui esta mañana la precampaña electoral se traslada al Congreso, La diputación permanente, el órgano de guardia de la Cámara, vota los seis decretos aprobados por el Gobierno tras convocar elecciones, Entre ellos el que amplía el permiso de paternidad y el que regula los alquileres, También hay sesión hoy en el parlament, Torra comparece por primera vez después de que el Tribunal Superior de Justicia de Cataluña le haya abierto una investigación por desobediencia por no retirar los lazos amarillos como exigió la Junta Electoral, En el juicio de Bankia, hoy declara el expresidente del BBVA, Francisco González, Rato le acusó de orquestar una conspiración con el gobierno para retirarle de la presidencia de la entidad,
F1: 1,388
F2: 0,097

FIGURE 5 Example of members of one population obtained with the optimization component

line, as well as the speed. The number of characters is automatically calculated by any programming language. For calculating the speed, the system assumes that the new subtitles must be shown in the same lapse of time available for the original ones, so the average speed is calculated by dividing the total number of characters to be processed by the total time available. With this average speed the time for each subtitle can be calculated by dividing the length of each line by the estimated average speed. These two parameters (size and presentation time) are the basis of the speed.

In this context, the speed is conditioned by the size. However, an optimum speed with a bad size value can be obtained. (For example a size of 2000 characters is unacceptable because the optimal number is 37 according the norms, but if they are shown in the screen for 134 seconds the speed would be optimal). Since the time on screen is established by the sound of the program, the aim of the algorithm is to determine the optimal subtitle size for the optimal time on screen, trying to get subtitles near to 37 characters and also to the optimal time for a speed of 15 characters per second as established by the Spanish regulation.

Therefore, the goal of the genetic algorithm is, given a set of words to be displayed together in the form of subtitles, to determine the optimum subtitle distribution maximizing the similarity to the optimal size and the optimal time on screen of an optimal speed. With these elements in mind, the algorithm has two objectives: size and time on screen (related to the speed). Nevertheless, despite being related, they are not always compatible, because long lines with a reasonable time can appear, although in any case a subtitle could last less than one second on the screen.

As mentioned at the beginning of this subsection, in order to address this multi-objective problem, the algorithm NSGA-II was selected. Accordingly, the two objective functions to be optimized are size and time on screen and are described as follows. The first function is related to the optimal size of the subtitle line. According to Spanish norms, the optimal size for a subtitle line is 37 characters. However, and due to the combination of words in each line, it is very difficult to obtain this optimal value in all the lines. Equation 2 represents the objective function for the size.

$$F_1 = \frac{1}{0,1 + |\overline{LinesLength} - OPTIMAL\_SIZE\| + \delta(LinesLength)} \tag{2}$$

Where:

- $\overline{\text{LinesLength}}$ is the average of the length of all the subtitles lines defined by the chromosome.
- OPTIMAL_SIZE is 37 (the optimal size of a line, defined by the norm).
- $\delta(\text{LinesLength})$ is the standard deviation of the length of the lines.

Since the less difference between the lengths of the lines and the optimal size the better (a minimization problem), it is necessary to convert the formula in order to focus on maximizing the result. Thus, the formula is represented as the inverse of this difference in order to obtain a function to be maximized. As this formula considers the average values, the standard deviation has been considered in order to favor uniformity in the lines, avoiding good average values based on combining long and short subtitled lines.

The second objective function is related to the time of the subtitle on the screen. The optimal speed is defined by the norm (15 characters per second), but, many times the speaker speaks extremely fast and this objective cannot be achieved maintaining the synchronization. Moreover and taking this into account, a subtitle cannot be on the screen for less than one second, because the human eye would not be able to read it. Considering all these elements, equation 3 represents the objective function for the time.

$$F_2 = \frac{1}{0,1 + |\overline{SubtitleTime} - OPTIMAL\_TIME\| + \delta(SubtitleTime)} \tag{3}$$

Where:

- $\overline{\text{SubtitleTime}}$ is the average of the time of all the subtitles defined by the chromosome.
- OPTIMAL_TIME is the optimal time for a subtitle with the optimal size (37 characters) and optimal speed (15 characters per second) according the Spanish norm. This time is 2.47 seconds.
- $\delta(\text{LinesLength})$ is the standard deviation of the time of the subtitles.
- If one of the SubtitleTime_i is less than 1 second, then the result of the function will be 0, because a subtitle cannot be on the screen for less than one second.

The Pareto approach has been applied for the selection of the individuals in the crossover process. A solution is the Pareto optimal when no other solution dominates it according to the objective functions. A solution "A" dominates the other solution "B" if "A" is equal to or better than "B" in all the objectives and better in at least one of the objectives. During the crossover process, the population is sorted based on this criterion in order to achieve a Pareto optimal solution.

The last generation of the algorithm should contain the best approach. The system will select the most adequate according to the time function, since the time is a key factor for reading the subtitles.

5 | EVALUATION

The persistence layer of the ATAD-SUB framework is a big data repository obtained from the integration of heterogeneous data from different sources. The data extracted from the DTT signal and the web services by the extraction component and processed by the pre-processing component feed this big data component. From this repository, the system calculates the speed at which the subtitles have been shown, determining the quality parameters, and associating them to information related to the type of program, which allows further analysis about, for example, the genre of the program (news, sports, entertainment, etc.). Finally, subtitles, speed data and information about the quality parameters or KPIs are sent to the genetic algorithm component in order to determine the best relationship between them. In the ATAD-SUB framework, the data processed and stored have been obtained within the time period between July 2012 and December 2017, that is, with more than 950 million records.

TABLE 1 GA configuration.

Parameter	Value
Population size	400
Generations	400
Crossover probability	70%
Mutation probability	3%
Number of runs	20

TABLE 2 Results obtained for subtitles analyzed during the evaluation process.

	Number	% Adequate Size	% Adequate Time	% Adequate Size and Time
Subtitles from DTT	3264	66,57	96,69	63,54
Subtitles from Genetic Algorithm (average)	3242,55	71,23	99,33	70,55

TABLE 3 Average values for objective functions of NSGA-II algorithm

Function	Average Value DTT	Average Deviation DTT	Average Value Genetic Algorithm	Average Deviation Genetic Algorithm
Time	1,15	1,21	2,01	1,23
Size	0,11	0,55	0,18	0,78

In order to define the evaluation scenario, different CBS have been created by merging different subtitles that have been shown consecutively on the original DTT broadcasting. For a more realistic testing scenario, different programs have been included in the CBS: deferred and semi-live (with live and deferred parts). To test the proposal, the authors have selected 3,264 subtitles for generating 665 CBS (560 for deferred programs and 105 for semi-live). The subtitles have been extracted from different programs and different DTT channels in order to cover the widest spectrum possible.

Table 1 shows the configuration of the GA. The genetic algorithm has been configured by trial and error. The system was tested using different sizes of CBS and different parameters for each size. The authors have selected the values that provided more reasonable results for the different sizes of CBS. The population size is 400. This number has been established by observation, because small CBS (blocks of text) required smaller population size to obtain adequate results, but larger CBS required higher population. This number has shown adequate results for both small and large CBS. The number of generations was also established at 400. As in the population size, small CBS required fewer generations to achieve a reasonable result, but larger CBS required more generations. This number provided good results in both cases. The crossover probability was established in 70% and the mutation probability was 3%. These values were also obtained by observation during the experiments. Finally, the algorithm has been executed 20 times for each CBS. The results shown in this section is the average values for all runs. In future research, the system could establish the parameters dynamically according to the size of the block in order to optimize process performance.

5.1 | Results

Table 2 shows the values obtained from the evaluation of the genetic algorithm with the subtitles extracted from the repository. As can be seen, the size of the generated subtitles is slightly different in the case of the genetic algorithm. It is also remarkable that the number of subtitles is different

TABLE 4 Results obtained for subtitles analyzed during the evaluation process for Semi-live subtitles.

	Number	% Adequate Size	% Adequate Time	% Adequate Size and Time
Subtitles from DTT	1764	55,67	97,05	53,06
Subtitles from Genetic Algorithm (average)	1728,8	54,20	99,64	53,84

TABLE 5 Average values for objective functions of NSGA-II algorithm for Semi-live subtitles

Function	Average Value DTT	Deviation Genetic Algorithm	Average Value Genetic Algorithm	Deviation Genetic Algorithm
Time	0,70	0,58	1,70	0,59
Size	0,06	0,01	0,10	0,03

TABLE 6 Results obtained for subtitles analyzed during the evaluation process for Deferred subtitles.

	Number	% Adequate Size	% Adequate Time	% Adequate Size and Time
Subtitles from DTT	1500	79,40	96,27	75,87
Subtitles from Genetic Algorithm (Average)	1513,76	90,66	98,98	89,64

TABLE 7 Average values for objective functions of NSGA-II algorithm for deferred subtitles

Function	Average Value DTT	Deviation DTT	Average Value Genetic Algorithm	Deviation Genetic Algorithm
Time	1,24	1,27	2,07	1,30
Size	0,12	0,60	0,20	0,14

in the case of the genetic algorithm, due to the fact that the genetic algorithm modifies the lines in order to adapt the size of the subtitles. However, despite the difference in the number of subtitles the percentage of most adequate size is not remarkably higher for the genetic algorithm. This can be due to the fact that the time available for the subtitles is not sufficient, and the genetic algorithm favors the time on screen of the subtitle. It also can be noted in the fact that the values of the time function for the genetic algorithm are better than the size ones when we compare with the values obtained with the DTT subtitles. However, the values of Table 3 show that the objective functions for Time (described in equation 3) and Size (described in equation 2) obtained by the genetic algorithm are higher than the values measured for the DTT subtitles. These functions represent the distance of the average time and size of the subtitles generated with respect to the most adequate established by Spanish norm. Despite the percentage of adequate size, the objective function values show that the size of the genetic algorithm are closer to the optimal values than the sizes of the evaluated DTT subtitles. This table includes the standard deviation of the results obtained by the Genetic Algorithm, and represents the average deviation in the runs of the experiments

Table 4 shows the values for semilive subtitles. Results shown are similar for both DTT and the optimized ones. Since all subtitles generated by the genetic algorithm are adequate from the point of view of the time, other subtitles have worst size value. Is is due to the fact that in the live parts of the program the speech is too fast and many times the time available for showing the subtitles is not enough. However, looking at the values of the evaluation functions in Table 5, the size of the subtitles generated by the genetic algorithm is closer to the optimal values. Table 6 shows the values for deferred subtitles. Results show that the percentages obtained by the genetic algorithm are higher than the measured in the DTT sample. Table 7 confirms that the values of the evaluation functions are higher in the case of the subtitles generated by the genetic algorithm.

It should be noted that the genetic algorithm is only considering two KPIs (time and size), but the genetic algorithm is not considering other criteria for generating the subtitles (like the punctuation criteria). For this reason, the genetic algorithm subtitles have been optimized for time and size, but they could lack other quality features that the DTT subtitles have. Future work will include new quality parameters in order to improve the results.

5.2 | Validation

In order to validate the proposal, in this research the authors have included a survey with real users to try to understand the relationship between speed, fidelity and understanding of the subtitles. For doing this, a Likert scale based survey has been defined. The Likert scale is the most comprehensive measurement tool in social science research surveys (Boone & Boone 2012). It measures the level of agreement or disagreement with an item or question allowing attitudes and levels of compliance to be measured. To do so, the first step is to make a series of affirmations or denials on the subject where the population surveyed must show their agreement or disagreement. Each item is given a score in order to later classify them

TABLE 8 Results obtained in the survey

	Video 1					Video 2					Video 3				
Question	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Q1	1		3	9	33	1	1	3	8	33	1		3	13	32
Q2		1	1	10	34		1	3	11	32		1	2	12	33
Q3		1	6	10	29	1	1	4	8	33		2	3	7	36
Q4			2	10	34	2		3	7	35		4	8	9	27

to obtain the results. The typical format of this type of measurements is with 5 levels of response: Strongly disagree, Disagree, Neither agree nor disagree, Agree, Totally agree.

In the validation test for this research, participants had to watch three videos subtitled at different speeds: 19, 14 and 9 characters per second. To achieve these speeds, the original subtitles were modified by adding words or summarizing the content of the subtitle (thus altering its fidelity). After seeing them, the participants had to indicate on a scale of 1 to 5 if:

- Q1: I can read the subtitles comfortably.
- Q2: I have had enough time to read all the subtitles
- Q3: I have had enough time to see the image and also to read the subtitles.
- Q4: I have understood the video

The test involved 141 people between 10 and 61 years old, some of them with partial hearing disability and others with total deafness. The results of the survey are depicted in Table 8. For each video, the answers to questions 1 to 4 obtained from the participants are indicated using the statistical mode of the answers for each category in the Likert scale (1 to 5).The accumulated of the answers from 141 participants for questions Q1 to Q4 and for each video is presented in Table 8. From the results obtained, it is observed that the users in general give maximum score (5) to all the questions for each video but the data are more dispersed in video 2 (with an average speed). Regarding the question Q3, the maximum value is obtained in video 3, the one with highest reading speed, being also the maximum number of the whole table.

The main conclusion obtained with the survey is that, regarding the three main KPIs (fidelity, speed and synchronization), the fidelity should not been modified because, if the contents are summarized or adapted, the comprehension of the users decay. Therefore, and taking this into account, the ACAD-SUB framework adapts the subtitle speed optimally considering quality parameters, and maintaining fidelity with the original subtitles.

5.3 | Discussion

Results show that the proposed framework is able to extract and process a huge number of subtitles from the DTT signal. These subtitles are extracted, classified, and processed to determine the program information and, finally, stored in order to be processed.

The framework also processes the data in order to allow further processes of the extracted contents. In this case, as part of a study of the quality of the subtitles, a genetic algorithm has been configured in order to suggest better subtitles according to different KPIs related to the size of the subtitles and time on screen for each subtitle.

In the proposed case of study, the framework selected blocks of text shown in the form of subtitles and proposed an alternative taking into account the size of the subtitles and the time on screen. As mentioned in the introduction, if the synchronization has to be preserved according to the users, the optimization should lay on the size of the subtitles and the time on screen. For allowing a correct reading, the number of characters must be around 37 and they should be on screen for a sufficient amount of time. It is known that the human eye cannot perceive elements shown on screen for less than one second, so the proposed functions try to maximize both time on screen and size of the subtitles. However, this is not always feasible, thus leading to the use of a multi-objective approach like the NSGA-II.

TABLE 9 Sample 1: good size and worst time

Subtitle Line	Text	Time	Size
1	Aqui esta mañana la precampaña electoral se	2,811310345	43
2	traslada al Congreso, La diputación permanente, el	3,268965517	50
3	órgano de guardia de la	1,503724138	23
4	Cámara, vota los seis decretos aprobados	2,615172414	40
5	por el Gobierno tras convocar elecciones,	2,680551724	41
6	Entre ellos el que amplía el permiso de	2,549793103	39
7	paternidad y el que regula los alquileres,	2,745931034	42
8	También hay sesión hoy en el	1,83062069	28
9	parlament, Torra	1,046068966	16
10	comparece por primera vez después de que el Tribunal	3,399724138	52
11	Superior de Justicia de Cataluña	2,092137931	32
12	le haya abierto una investigación	2,157517241	33
13	por desobediencia por no retirar	2,092137931	32
14	los lazos amarillos como exigió la Junta Electoral, En	3,530482759	54
15	el juicio de Bankia, hoy declara el	2,288275862	35
16	expresidente del BBVA, Francisco González,	2,745931034	42
17	Rato le acusó de orquestar una	1,96137931	30
18	conspiración con el gobierno para	2,157517241	33
19	retirarle de la presidencia de la entidad,	2,745931034	42
	Total time	46,22317241	
	Funtion 1 (Size)	0,103090526	
	Function 2 (Time)	1,337328701	

Tables 9 and 10 show the lines of subtitles generated by two candidate chromosomes of the final population in one of the tests. The values of the functions are very close in both cases and the size of the lines produced by the two chromosomes is almost the same. However, there is one difference: the chromosome represented in Table 9 has 19 lines and the chromosome in Table 10 has 18. Let us analyze why. On the one hand, Table 10 presents a time function 3 value higher than Table 9, which means that the time of the lines of subtitles of Table 10 are nearer to the optimal than the lines shown in the other table. On the other hand, Table 9 presents a function size 2 value higher than the one obtained for Table 10, which means that the size of the lines of Table 9 are nearer to the optimal values. Analyzing both samples, it can be seen that the difference between them are in lines 8 and 9. In Table 9, line 9 has 16 characters (apparently good but far from the optimal 37) and the time on screen is 1.04 seconds (almost at the limit of perception). However, in Table 10, line 9 has 52 characters (more than 37) but the time on the screen is higher (around 3.4 seconds). It means that the line has more characters but they are on screen longer so the probability of being read is higher. It also can be seen that line 9 of Table 10 is the junction of lines 8 and 9 of Table 9. (That is the reason why both chromosomes have a different number of lines and produce different measures.) Thus, the algorithm has generated two possibilities, one is better in terms of size and the other one is better in terms of time. This is an example of how time and size can be in conflict and how they should be considered as separated objectives.

This experiment has only considered two of the possible KPIs. However, other criteria can be taken into account, such as rules relative to punctuation or good practices for separating the lines. Future work will consider more objectives in order to improve the quality of the subtitles, considering the restrictions of synchronization and fidelity.

6 | CONCLUSIONS

In Spain, the subtitling service on television for the deaf has been improving in quantity since the General Law on Audiovisual Communication was enacted in 2010. This law establishes a series of quality standards that must be followed in the subtitling process. In addition, according to the state of art included in this paper, the subtitles in audiovisual content broadcast by DTT are a challenge from the point of view of quality, especially in live

TABLE 10 Sample 2: good size and better time

Subtitle Line	Text	Time	Size
1	Aqui esta mañana la precampaña electoral se	2,811310345	43
2	traslada al Congreso, La diputación permanente, el	3,268965517	50
3	órgano de guardia de la	1,503724138	23
4	Cámara, vota los seis decretos aprobados	2,615172414	40
5	por el Gobierno tras convocar elecciones,	2,680551724	41
6	Entre ellos el que amplía el permiso de	2,549793103	39
7	paternidad y el que regula los alquileres,	2,745931034	42
8	También hay sesión hoy en el parlament, Torra	2,942068966	45
9	comparece por primera vez después de que el Tribunal	3,399724138	52
10	Superior de Justicia de Cataluña	2,092137931	32
11	le haya abierto una investigación	2,157517241	33
12	por desobediencia por no retirar	2,092137931	32
13	los lazos amarillos como exigió la Junta Electoral, En	3,530482759	54
14	el juicio de Bankia, hoy declara el	2,288275862	35
15	expresidente del BBVA, Francisco González,	2,745931034	42
16	Rato le acusó de orquestar una	1,96137931	30
17	conspiración con el gobierno para	2,157517241	33
18	retirarle de la presidencia de la entidad,	2,745931034	42
	Total time	46,28855172	
	Function 1 (Size)	0,096885709	
	Function 2 (Time)	1,387658315	

broadcasting. In this sense, there are some quality parameters or indicators (also known as Key Performance Indicator or KPI) that should be taken into account for showing subtitles that are readable and understandable. Within the available set of KPIs, there are three that are of importance for this purpose and sometimes are in conflict: synchronization, speed and literal subtitling (fidelity or accuracy).

This article presents the ATAD-SUB framework for the integration and processing of heterogeneous information associated with the subtitling of audio-visual content from different sources. Moreover, the framework provides an automatic adjustment of the speed of subtitling in broadcasting regarding quality indicators by means of a genetic algorithm approach. In this research, the authors have worked with the data obtained within the time period between July 2012 and December 2017, that is, with more than 950 million records.

Different scenarios have been defined to test the proposal: deferred, semi-live (with live and deferred parts) and live programs. In these scenarios, the optimal speed for subtitles have been found considering the quality parameters of the KPIs. Next, the results obtained have been validated by final users with hearing impairment by means of questionnaires based on the Likert scale. In all of the scenarios, the framework is able to gather and process heterogeneous information and return as output the optimal time and size of the subtitles regarding quality parameters by means of the multi-objective genetic algorithm NSGA-II.

In this sense, the results obtained in the experimentation show that the proposed framework is able to extract and process a huge number of subtitles from the DTT signal. These subtitles are extracted, classified, and processed to determine the program information and, finally, stored in order to be processed. The framework also processes the data in order to allow further processes of the extracted contents. In this case, as part of a study of the quality of the subtitles, a genetic algorithm has been configured in order to suggest better subtitles according to different KPIs related to the size of the subtitles and time on screen for each subtitle The results indicate that the system is able to estimate the best relationship between the speed of the subtitles and maintaining the quality levels defined for this research.

With regard to the validation performed by experts and users of this domain, the main conclusion of the survey is that considering the three main KPIs (fidelity, speed and synchronization), the fidelity should not be modified because, if the contents are summarized or adapted, the comprehension of the users decay. Therefore, and taking this into account, the ATAD-SUB framework adapts the subtitle optimally (considering time and size) considering quality parameters and maintaining fidelity with the original subtitles.

One of the most relevant aspects of subtitle quality is the speed at which they are shown on the screen, due to the fact that a too high speed (less time on screen) will make them difficult to read and the information hard to understand. In order to determine whether the speed at which the subtitles are being shown is adequate, firstly, it is necessary to process all the information associated with the broadcast of the digital TV channels including data from different sources. Accordingly, the idea behind this research is tackling the main technical problem faced by accessible television for overcoming communication barriers that affect a large number of people, thereby generating an inclusive solution for all. Thus, the expected contribution of the research is a solution for improving the quality of life of people with disabilities when consuming audiovisual contents. At the same time, there are other users who can directly benefit from this system: children, the elderly, persons with intellectual disabilities, people learning languages, since the possibility of improving reading and writing skills is an added value to subtitling services. In addition, subtitling solutions like the one presented here, can also benefit the general public in certain noisy environments (public transport, commercial areas, etc.), and their knowledge could sensitize and prepare them for possible age-associated hearing loss. Moreover, using this technology for distribution in television or streaming on the Internet can be a key asset for the education sector.

As a future research line, different improvements in the big data infrastructure can be performed for a better distribution of data in the nodes (taking advantage of this kind of solutions for a better performance, fault tolerance and scalability). Moreover, for improving the efficiency of the framework and to increase data analysis operations a new implementation for processing and generating big data sets with a parallel, distributed algorithm on a cluster could be included (as the MapReduce programming model for processing and generating big data sets with a parallel, distributed algorithm on a cluster). Furthermore, as future improvements in the optimization component, different possibilities can be analyzed to be included in the objective functions of the NSGA-II algorithm (taking other approaches found in the literature review into consideration). In addition, different KPIs such as color and labels can be tested in the objective functions or trying, instead of Pareto, another function to determine which are more suitable, pondering factors such as size or line length. Finally, taking advantage of the modularity of the framework and considering further synergies, ATAD-SUB could be integrated with other systems receiving the transcription of an audiovisual content obtained by means of an Automatic Speech Recognition component (ASR). The ATAD-SUB framework will obtain the correct time and size of the subtitles allowing a full automatic system for the subtitling process. Moreover, the framework's adaptation to languages other than Spanish is currently being evaluated.

References

AENOR. (2011). *UNE 133300:2011. Information on content for Digital Terrestrial Television Broadcasting* (Tech. Rep.). UNE. Retrieved from `https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0048560`

AENOR. (2012). *UNE 153010:2012. Subtitulado para personas sordas y personas con discapacidad auditiva* (Tech. Rep.). UNE. Retrieved from `https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0049426`

AMADIS. (2016). *IV Congreso de accesibilidad a los medios audiovisuales para personas con discapacidad* (Tech. Rep.). Retrieved from `http://www.rpd.es`

Ameixa, D., Coheur, L., & Redol, R. A. (2013). *From subtitles to human interactions: introducing the subtle corpus* (Tech. Rep.). Tech. rep., INESC-ID (November 2014).

Ando, A., Imai, T., Kobayashi, A., Homma, S., Goto, J., Seiyama, N., … others (2003). Simultaneous subtitling system for broadcast news programs with a speech recognizer. *IEICE TRANSACTIONS on Information and Systems*, *86*(1), 15–25.

Baggett, P. (1989). 6 understanding visual and verbal messages. In *Advances in psychology* (Vol. 58, pp. 101–124). Elsevier.

Bhogal, J., & Choksi, I. (2015). Handling big data using nosql. In *2015 ieee 29th international conference on advanced information networking and applications workshops* (pp. 393–398).

BOE. (2010). *Ley 7/2010 general de la comunicación audiovisual* (Tech. Rep.). Madrid: Author. Retrieved from `https://www.boe.es/buscar/act.php?id=BOE-A-2010-5292`

Boone, H. N., & Boone, D. A. (2012). Analyzing likert data. *Journal of extension*, *50*(2), 1–5.

Brewer, E. A. (2000). Towards robust distributed systems. In *Podc* (Vol. 7).

Camacho, D., Lara-Cabrera, R., Merelo-Guervós, J., Castillo, P. A., Cotta, C., Fernández-Leiva, A. J., … Chávez, F. (2018). From ephemeral computing to deep bioinspired algorithms: New trends and applications. *Future Generation Computer Systems*, *88*, 735–746.

Carpuat, M., Vyas, Y., & Niu, X. (2017). Detecting cross-lingual semantic divergence for neural machine translation. In *Proceedings of the first workshop on neural machine translation* (pp. 69–79).

Chodorow, K. (2013). *Mongodb: the definitive guide: powerful and scalable data storage.* " O'Reilly Media, Inc.".

Corbellini, A., Mateos, C., Zunino, A., Godoy, D., & Schiaffino, S. (2017). Persisting big-data: The nosql landscape. *Information Systems*, *63*, 1–23.

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T. (2000). A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii. In M. Schoenauer et al. (Eds.), *Parallel problem solving from nature ppsn vi* (pp. 849–858). Berlin, Heidelberg: Springer Berlin Heidelberg.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, *6*(2), 182–197.

de Castro, M., Carrero, D., Puente, L., & Ruiz, B. (2011). Real-time subtitle synchronization in live television programs. In *2011 ieee international symposium on broadband multimedia systems and broadcasting (bmsb)* (pp. 1–6).

De Linde, Z., & Kay, N. (2016). *The semiotics of subtitling.* Routledge.

Del Ser, J., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., … Herrera, F. (2019). Bio-inspired computation: Where we stand and what's next. *Swarm and Evolutionary Computation*, *48*, 220–250.

ETSI. (2016). *ETSI EN 300 468 V1. 15.1. Digital Video Broadcasting (DVB). Specification for Service Information (SI) in DVB systems* (Tech. Rep.). Retrieved from https://www.etsi.org/deliver/etsi_en/300400_300499/300468/01.15.01_60/en_300468v011501p.pdf

Ferriol, J. L. M. (2013a). Subtitle reading speed: A new tool for its estimation. *Babel*, *59*(4), 406–420.

Ferriol, J. L. M. (2013b). Subtitle reading speeds in different languages: the case of lethal weapon. *Quaderns: Revista de traducció*(20), 201–210.

Fors-Isalguez, Y., Hermosillo-Valadez, J., & Montes-y Gómez, M. (2018). Query-oriented text summarization based on multiobjective evolutionary algorithms and word embeddings. *Journal of Intelligent & Fuzzy Systems*(Preprint), 1–10.

Gambier, Y. (2003). *Introduction: Screen transadaptation: Perception and reception.* Taylor & Francis.

Gao, J., Zhao, Q., Li, T., & Yan, Y. (2009). Simultaneous synchronization of text and speech for broadcast news subtitling. In *International symposium on neural networks* (pp. 576–585).

Garcia, J. E., Ortega, A., Lleida, E., Lozano, T., Bernues, E., & Sanchez, D. (2009). Audio and text synchronization for tv news subtitling based on automatic speech recognition. In *2009 ieee international symposium on broadband multimedia systems and broadcasting* (pp. 1–6).

Ghia, E. (2012). The impact of translation strategies on subtitle reading. *Eye tracking in audiovisual translation*, 157–182.

González-Carrasco, I., Puente, L., Ruiz-Mezcua, B., & López-Cuadrado, J. L. (2019). Sub-sync: Automatic synchronization of subtitles in the broadcasting of true live programs in spanish. *IEEE Access*, *7*, 60968-60983. doi: 10.1109/ACCESS.2019.2915581

Gottlieb, H. (2012). Subtitles–readable dialogue? *Eye tracking in audiovisual translation*, 37–82.

Grefenstette, J. J. (2013). *Genetic algorithms and their applications: proceedings of the second international conference on genetic algorithms.* Psychology Press.

Huang, Y., Hsu, L., & Sandnes, F. (2007, April). An intelligent subtitle detection model for locating television commercials. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *37*(2), 485-492. doi: 10.1109/TSMCB.2006.883428

Jensema, C. (1998). Viewer reaction to different television captioning speeds. *American annals of the deaf*, *143*(4), 318–324.

Kar, A. K. (2016). Bio inspired computing–a review of algorithms and scope of applications. *Expert Systems with Applications*, *59*, 20–32.

Kundu, A., Das, D., & Bandyopadhyay, S. (2013). Scene boundary detection from movie dialogue: A genetic algorithm approach. *Polibits*(47), 55–60.

Lamprier, S., Amghar, T., Levrat, B., & Saubion, F. (2007). Seggen: A genetic algorithm for linear text segmentation. In *Ijcai* (Vol. 7, pp. 1647–1652).

Lertwongkhanakool, N., Punyabukkana, P., & Suchato, A. (2013). Real-time synchronization of live speech with its transcription. In *2013 10th international conference on electrical engineering/electronics, computer, telecommunications and information technology* (pp. 1–5).

Müller, M., & Volk, M. (2013). Statistical machine translation of subtitles: From opensubtitles to ted. In *Language processing and knowledge in the web* (pp. 132–138). Springer.

Ofcom. (2005). *Subtitling–an issue of speed?* Office of Communications London.

Parmenter, D. (2015). *Key performance indicators: developing, implementing, and using winning kpis.* John Wiley & Sons.

Pražák, A., Psutka, J. V., Psutka, J., & Loose, Z. (2013). Towards live subtitling of tv ice-hockey commentary. In *2013 international conference on signal processing and multimedia applications (sigmap)* (pp. 151–155).

Romero-Fresco, P. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics*, *6*.

Romero-Fresco, P. (2010). Standing on quicksand: hearing viewers' comprehension and reading patterns of respoken subtitles for the news. In *New insights into audiovisual translation and media accessibility* (pp. 175–194). Brill Rodopi.

Romero-Fresco, P., & Pérez, J. M. (2015). Accuracy rate in live subtitling: The ner model. In *Audiovisual translation in a global context* (pp. 28–50). Springer.

Ruiz, B., Quintana, I., García Crespo, Á., de Castro, M., Souto, M., González, I., … Sánchez Pena, J. M. (2014). Guía de accesibilidad al teatro a través del subtitulado y la audiodescripción.

Sahoo, P., Ekbal, A., Saha, S., Mollá, D., & Nandan, K. (2016). Semi-supervised clustering of medical text. In *Proceedings of the clinical natural language processing workshop (clinicalnlp)* (pp. 23–31).

Sanchez-Gomez, J. M., Vega-Rodríguez, M. A., & Pérez, C. J. (2018). Extractive multi-document text summarization using a multi-objective artificial bee colony optimization approach. *Knowledge-Based Systems*, *159*, 1–8.

Simpson, M., Barrett, J., Bell, P., & Renals, S. (2016). Just-in-time prepared captioning for live transmissions.

Song, W., Li, C. H., & Park, S. C. (2009). Genetic algorithm for text clustering using ontology and evaluating the validity of various semantic similarity measures. *Expert Systems with Applications*, *36*(5), 9095–9104.

Song, W., & Park, S. C. (2009). Genetic algorithm for text clustering based on latent semantic indexing. *Computers & Mathematics with Applications*, *57*(11-12), 1901–1907.

Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. *Knowledge-Based Systems*, *24*(7), 1024–1032.

Utray, F., de Castro, M., Moreno, L., & Ruiz-Mezcua, B. (2012, may). Monitoring Accessibility Services in Digital Television. *International Journal of Digital Multimedia Broadcasting*, *2012*, 1–9. Retrieved from `http://www.hindawi.com/journals/ijdmb/2012/294219/` doi: 10.1155/2012/294219

Verma, A., Llorà, X., Goldberg, D. E., & Campbell, R. H. (2009). Scaling genetic algorithms using mapreduce. In *2009 ninth international conference on intelligent systems design and applications* (pp. 13–18).

□