Innovative Applications of Natural Language Processing and Digital Media in Theatre and Performing Arts

Charalampos Tsioustas
Innovative Technologies Centre, Greece
Daphne Petratou
Innovative Technologies Centre, Greece
Maximos Kaliakatsos-Papakostas
Institute for Language & Speech Processing, Athena Research & Innovation Centre, Greece
Vassilis Katsouros
Institute for Language & Speech Processing, Athena Research & Innovation Centre, Greece
Apostolos Kastritsis
Institute for Language & Speech Processing, Athena Research & Innovation Centre, Greece
Konstantinos Christantonis
Dept. Information & Electronic Eng., International Hellenic University, Greece
Konstantinos Diamantaras
Dept. Information & Electronic Eng., International Hellenic University, Greece
Michael Loupis
National and Kapodistrian University of Athens, Greece

Abstract

The objective of our research is to investigate new digital techniques and tools, offering the audience innovative, attractive, enhanced and accessible experiences. The project focuses on performing arts, particularly theatre, aiming at designing, implementing, experimenting and evaluating technologies and tools that expand the semiotic code of a performance by offering new opportunities and aesthetic means in stage art and by introducing parallel accessible narrative flows. In our novel paradigm, modern technologies emphasize the stage elements providing a multilevel, intense and immersive theatrical experience. Moreover, lighting, video projections, audio clips and digital characters are incorporated, bringing unique aesthetic features. We also attempt to remove sensory and language barriers faced by some audiences. Accessibility features consist of subtitles, sign language and audio description. The project emphasises on natural language processing technologies, embedded communication and multimodal interaction to monitor automatically the time flow of a performance. Based on this, pre-designed and directed stage elements are being mapped to appropriate parts of the script and activated automatically by using the virtual "world" and appropriate sensors, while accessibility flows are dynamically synchronized with the stage action. The tools above are currently adapted within two experimental theatrical plays for validation purposes.

Keywords: Natural Language Processing, Digital Art, Virtual Characters, Immersive Theatre
JEL classification: C63

Acknowledgments: This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (NLP-Theatre, project code: T1EDK-00508).

Paper type: Research article
Received: May 22, 2020
Accepted: Jun 9, 2020

Introduction

There is a wide and diverse compilation of elements that constitute a theatrical production. Designing and integrating them in the performance is a very challenging, interdisciplinary process. In our project, we address the challenge of synchronising a number of digital performance enhancements, such as visual projections, audio media, sign language media and supertitles. Live captioning, to start with, is considered often an accessory, which provides spectators minimal assistance in understanding the dialogues (Thonon, 2018). However, cultural life, especially in multi-cultural cities, has set new standards in the field of art accessibility, which renders live captioning increasingly essential (Thonon, 2018). This also applies to a variety of stage elements and techniques that create interesting experiences, while contributing to the performances' accessibility (Secară, 2018). Furthermore, a whole sector has developed within performing arts, dealing with these elements; it encompasses fields such as technical management, acoustics, computer programming, translation and scenography (Thonon, 2018). Although this sector is developing in a rapid pace, and theatres around the world make admirable steps towards art accessibility, one issue that remains is the lack of a complete, applicable solution towards stage element automation and synchronisation; technicians operate manually stage elements, as the performance takes place, while important practical difficulties (e.g. interpreter fatigue, low-visibility, or artistic requirements) inhibit the use of sign language interpretation media. Overall, it is very difficult for smaller art groups and theatres to access the necessary equipment, software and technical knowledge in order to explore the possibilities technology has to offer. One way to address these issues is natural language processing; speech recognition and voice generation alongside machine learning techniques are already applied in improvisational theatre (Henry, 2019). In such applications, artificial improvisors serve as chatbots or artificial conversational agents and they can fully participate in natural human conversations (Mathewson & Mirowski, 2017), while humanoid robots can communicate via verbal and non-verbal interactions with humans (Perkowski et al., 2005).

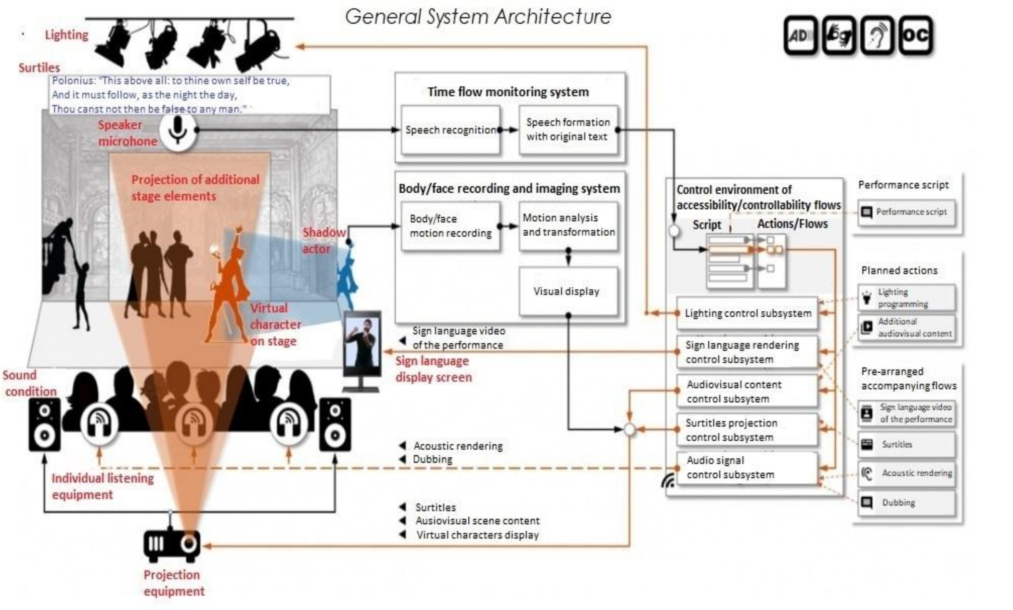
As far as the technological scene augmentation is concerned, over time, various interesting attempts have been made. Technology enters the scene in a variety of ways, transforming it and creating hybrids such as "digital drama" (Shaw, 2012). Virtual characters interact on stage with actors, video projections are combined with traditional stage elements creating innumerable possibilities for the artists. In a recent production of 'The Tempest', Ariel's character appeared on stage as a digital entity, while controlled by a real actor (Fiveash, 2016). Similarly, the performance Le Sacre du Printemps, combined stage performance with live music and real-time 3D graphics (Ars Electronica, 2006). Using projections and video mapping of computer graphics can create an interesting artistic language. These graphics are quite different depending on the surface on which they are displayed and the appropriate synchronization with the flow of the performance creates the impression of their interaction with other audio-visual material and with the characters themselves. An example is the multi-award winning play "Sunday in the Park with George" (Bird, 2006) and the performance “Hakanai” (Siddiqui, 2015).

Finally, yet importantly, the technique of an Arduino controlling stage lighting has been recently explored (Heimbach et al., 2016). The Arduino acts as a controller and communicates in a DMX512 environment, which is of particular interest, due to the need for flawless automation. This approach could replace the usual digital/analogue lighting console.

Methodology

In the heart of our research lies a system for monitoring the time flow of the performance. It is an advanced automatic speech recognition system (ASR) that monitors the progress of the performance by comparing the current dialogues of the actors with the dramatic text of the performance. The system can be sufficiently robust to be able to analyse the stage speech, effectively addressing the challenges as well as the technical limitations of acoustics and sound recording.

Figure 1
General System Architecture

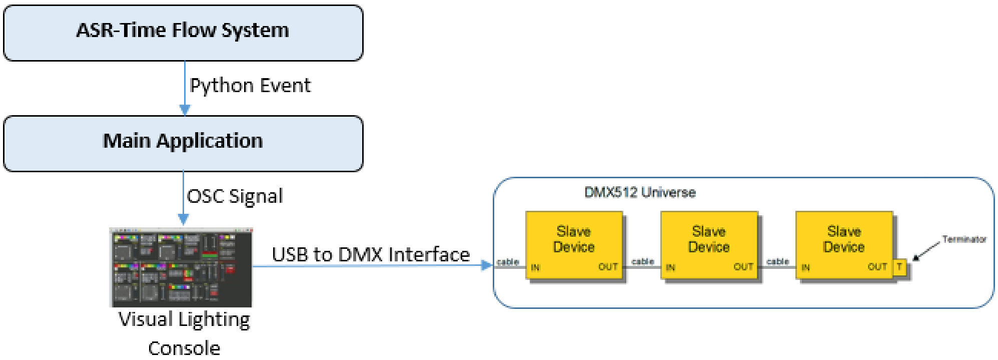


Source: Author’s Illustration

As shown in Figure 1, the supertitles are synchronized automatically during the performance, ensuring the precise display - in the viewer's language - and avoiding the confusion caused by time related misalignments between stage action and live captioning. Stage lighting and additional audio-visual material, which is usually scheduled based on the script, is synchronized automatically, or even dynamically, following the real flow of the stage action. The floor technician is no longer required to be alert in order to handle the lighting through an analogue or digital console and the relative time of each queue can be automatically triggered by the central system, which in turn enables via OSC signals the digital console to send a DMX signal to the appropriate DMX lights, Figure 2.

One of the main advantages of this system is the introduction of modern elements and aesthetic codes. These elements also concern digital/virtual characters, which can be included and participate in a performance, contributing aesthetic features and stage points, beyond those possible by human actors. These virtual characters are the result of digital processing, able to have a variety of forms depending on their role and included in the scene through video projection. The virtual characters perform their role either independently, based on a predetermined script, or driven in real time by a natural actor, acting behind the scenes. The movement, gestures and facial expressions of the shadow actor can be transferred in real time to the virtual character on stage.

Figure 2
Lighting Control Subsystem



Source: Author's Illustration, DMX 512 Universe (2013)

A special case is also the synchronised translation of the performance in sign language. The video recording of the semantic performance is done during the preparation stage while, during the performance, this is displayed in sync at a selected point of the scene or on its margin. Furthermore, the possibility of dubbing a performance is of particular interest. In this case, the viewers watch a performance in its original language while at the same time, with a simple audio aid, they can listen to the dialogues and the narration in their own language. Similarly, the compilation is done at the point of preparation of the performance and it is dynamically synchronized with the real time development of the action by the flow monitoring system. The "life cycle" of the system incorporates three distinct phases:

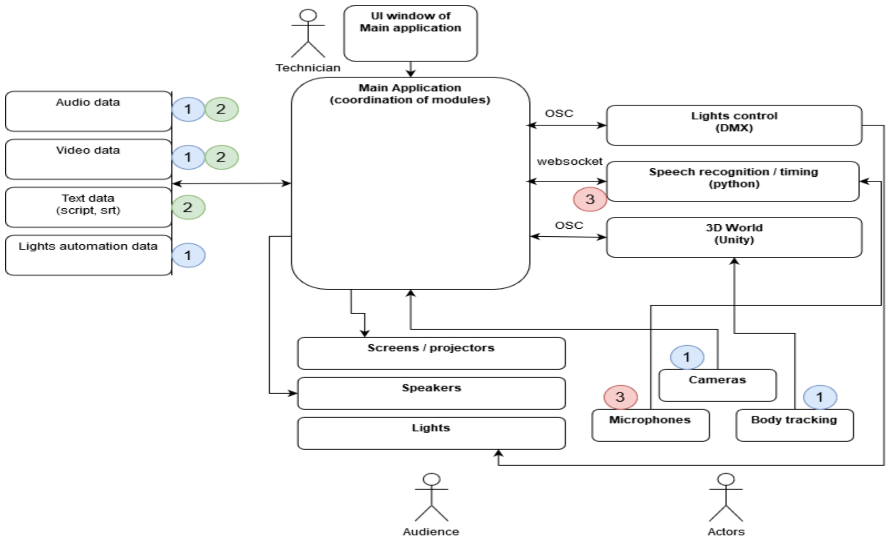
1. Rehearsal setup and recording.
2. Rehearsal processing and content completion.
3. Automation-assisted performance.

A functional overview of the system is shown in Figure 3, where numbers adjacent to components indicate which parts are employed at which phase of the life cycle, or what content is generated. The role of each phase is analysed as follows:

Phase 1: Rehearsal setup and recording phase

This phase sets the basis for developing key-content, while it also provides for a rough timing reference for the modules that are responsible for automating the discussed processes. Initially, the technician sets up all the events of the performance (3D world, lights, audio and video events) by adding events in the custom Main App that is being developed in the context of the project. When the rehearsal begins, the technician manually activates the events based on the script, using the User Interface (UI) of the Main App. During the rehearsal, the lights, audio/video and 3D world performance are connected with the Main App, allowing direct activation of these modules whenever the technician decides to activate them, based on the script. There are two key elements at the rehearsal setup and recording phase. (a) When the rehearsal begins, the technician starts the global timer of the Main App. When events are activated through the Main App during the rehearsal, they are assigned with a timestamp according to the global timer of the Main App. (b) The rehearsal is also recorded with a conventional camera (video and audio).

Figure 3
Overview of the system



Source: Author's Illustration

Phase 2: Rehearsal processing and content completion phase

From the previous phase the technician has obtained the reference recording (audio/video) and a set of rough activation times for the rehearsal events, aligned with the reference recording. At the second phase, the material obtained from the first phase is processed in two ways: (a) The activation times of the events in the rehearsal are fine-tuned according to the reference video of the recording (to be adjusted for potential inaccuracies that could have occurred due to the human factor during the rehearsal). (b) New content is created based on the reference video for completing the functionality of all modules of the system. Specifically, based on the reference recording: supertitle files are generated (.srt format); audio files are recorded that include dubbing in multiple languages as well as audio descriptions; and videos in sign language are prepared.

During the second phase, after the creation of additional content is completed, the new files are inserted as timestamped events in the existing list of events that have been prepared and activated during the first phase. At this point, the technician has all the necessary material (in the form of timestamped events) to reproduce the rehearsal (through the reference video/audio) with all the content, including supertitles, audio/video files, 3D world and light events. There is no automation involved in this phase, however. All events and material are adjusted to fit perfectly to the reference recording. Therefore, the technician can, at any point, playback the video of the rehearsal and observe how the automations align, finetuning the timing of events whenever necessary. The *main assumption* for developing the system is that it is possible to accurately and quickly identify which phrase is being spoken by the actors any time during the performance (after the rehearsal) using ASR. A consequence of this is that the system can accurately identify the specific supertitle corresponding to the actors' speech at any given time. Based on this, the system "anchors" all time stamped events (the entire content of the rehearsal) to their *closest preceding supertitle*. The timestamp of each event is not any more absolute, according to the global time of the reference rehearsal recording: the timestamp of each event becomes relative to the timestamp of its closest preceding supertitle.

Phase 3: Automation-assisted performance

This phase is the actual performance, where events can be activated automatically based on when their closest preceding supertitle is identified from the ASR system and their time distance from the current supertitle. The technician, however, is given the option to select which events should *not* be automated, since there are events that depend on extra-lingual conditions; e.g. lightning could strike (activation of audio and lights events) when an actor looks at the sky without speaking. At this phase, the technician can declare the system in "panic mode", if something goes wrong, by pressing a button. If this happens, the system switches off and hides all events not present in phase 1, effectively switching back to the "rehearsal mode", where the technician should feel comfortable to take full control of all the events setup during phase 1. At any point, the system can return to normal performance mode (ending the "panic mode") continuing from the next supertitle that the ASR system identifies.

Digital scenography and avatars

The basic outline of our architecture concerning a live performance with digital scenography and avatars driven by actors can be summarized in three basic stages: the creation of visual content (virtual characters, world space and digital scenography), motion capturing, and live performance/ visualisation/interaction. The creation of the virtual characters is a tedious process that involves various stages, working as a pipeline. Modelling in 3D is the process of creating and developing a mathematical representation of a surface of an object in three dimensions and is quite similar to plastic arts such as sculpting. These three-dimensional models can be projected into a two-dimensional image through a process that is called rendering. Sequentially, the process of applying textures (an array of colour pixels that consist of an image) to parts of the model is called texturing process and is quite important if we aim at achieving a photorealistic quality. In our case, in order to minimize the effort of producing avatars for our theatrical performances, we have used MakeHuman, an open source software, for modelling and initial texturing. The textures that are created can further be processed to create images with different levels of detail (displacement or normal maps etc) that apply as layers of increasing complexity and define how light will behave on those surfaces.

Creation of world space and digital scenography. Digital scenography is the application of new and emerging technologies to the centuries old practice of scenography that define space, atmosphere and mood. 'Mise en scene' is a term usually used in cinema and can be used to set the "mood" of the narration, provide a certain perspective upon a scene and create a specific atmosphere. The relation between the objects and the world, the use of lighting or the way the camera is placed, can provide a non-linguistic arrangement and presentation of what is going to happen and why. In our case we use Unity Game Engine to construct a virtual world that will be projected on stage. Depending on the scenario we can choose elements of the narrative that can be augmented so as to serve the purpose of the storytelling in such a way that every element in the scene, becomes a part of the narration thus stimulating the audience's emotional involvement.

Motion capture is the process of recording a live motion event and translating it in usable mathematical terms. When motion capture systems are used in live theatrical performance, visualisation technology needs to be combined with a tracking system that considers head and body movements as inputs to the digital world. An actor should be able to perform his acting in terms of body movement as naturally as possible, so we must be careful on the motion capture technique to be used. In our case we have chosen to experiment with an RGB sensor (Kinect sensor) for the body movement so as not to limit the actor's body with wearable equipment. Despite the fact that RGB sensors have some limitations, it is an inexpensive way to track body movements and later transfer them into the virtual world. We also experimented with Faceware helmet camera so as to capture the actor's facial expressions.

The final step in our implementation involves retargeting, the procedure of transferring a motion from the actor to the digital avatar. Most virtual character's animations are based on armatures to manipulate the character's body parts and the creation of this armature is a process called rigging. Armatures behave as the character's skeleton with a set of bone segments and their constraints chained together. Each bone of the skeleton is associated with a well-defined set of vertices defining the character's mesh (skinning). This structure allows animators to control the movements and deformations of the character. In our case, we need to map the rotations and translations of the actor's body parts to the specific bones and joints of the avatar. Consequently, we need to capture the facial expressions as well and integrate it with the corresponding body movements. The avatar is placed in the scene inside the virtual world with all the elements that define the virtual 'mise en scene'. The projection of the virtual world and merging it with the physical space creates a mixed environment where narrative, empathy and immersion work in an interchangeable way. Exploring the possibilities of user's interaction inside or outside the scenario aims at producing added value that hopefully augments the performance and its dramaturgy.

Results

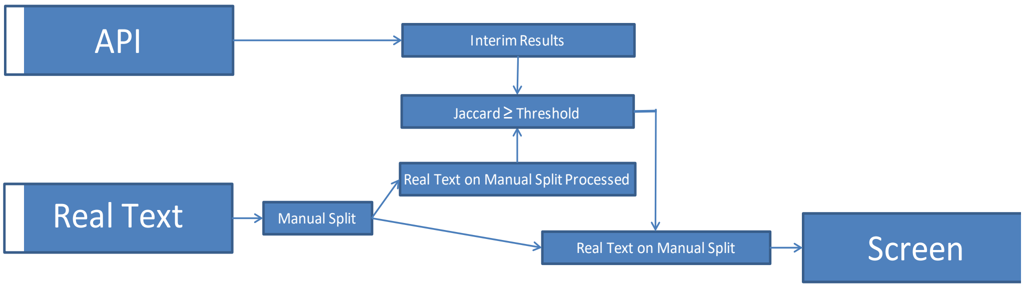
In this section we analyse the tools and methods used to achieve the live captioning on the theatrical play and the special needs and requirements that our system should meet. In contrast to films and TV series, in theatre it is common for actors to improvise and skip dialogues. For that reason, a pre-made queuing list, may jeopardize the real flow of dialogues. It is very unlikely to record two consecutive performances without any deviations. Many efforts so far on live captioning regard different scenarios, obviously, with different restrictions (Gao et. al., 2011, González-Carrasco et al., 2019). In general, live captioning brings a small delay on the classic way of subtitling. Usually, the audio is being processed and the corresponding text comes partially into view. Until now, this was manually done by someone who listened to the audio live and tried to match the text. In our case, we use the Google Speech-To-Text service. The Google's API is considered one of the best in 2020 and the service is available for a plethora of languages, including Greek. The Speech to Text service is mostly efficient and accurate when certain factors are ensured, e.g. that the audio is captured with a sampling rate of 16 kHz or higher. In addition, if noise is present (which is unavoidable in theatrical plays) the microphone should be as close to the speaker as possible. The recognizer is designed to ignore background voices and noise without additional noise-cancelling. Lastly, the audio captioning in case of multiple speakers should be either sent separately, if each person is recorded on a separate channel, or mixed in a single channel for all speakers. Multiple people talking at the same time or at different volumes may be interpreted as background noise and ignored.

When all the factors above are ensured the accuracy of such systems is impressive. However, they are not able to reach an absolute result yet. Thus, they are not reliable as standalone systems on live captioning and a second system is essential to transform their results into the final captions. In our case the real text of the play is available and based on that, we align the API responses. The API returns two types of results, the interim and the final. In general, while the user keeps speaking, the service returns interim results as *strings* with a level of confidence. The interim results may differ significantly from each other, especially at the beginning. The final result is only available when the system detects silence for approximately more than a second. The final result has the highest confidence and is the most reliable; nevertheless, its delay introduces certain difficulties for a synchronization system. Using the real text of our experimental play (The Little Prince by Antoine de Saint-Exupéry), results showed that the API managed to transcribe our audio into text without any mistakes.

In the case of an immediate response by a second actor the prerequisite silence gap does not exist and therefore the system keeps transcribing the audio without separating the speakers. The major challenge here is to manage the flow of results returned by the API, while matching it with the real text. Our system, however, should be able to prevent some common errors in a theatrical play. As already mentioned, actors may accidentally miss the flow of real text and skip or improvise lines. Therefore, the captions should not include sub-phrases that the actor missed. Therefore, the first constraint in our system is that the real text should be split into smaller phrases by dividing regular and minor sentences. Moreover, interjections stand as normal text and are not being processed. A second characteristic is that the API does not return punctuation and therefore misses the proper mark of capital letters. For the part of texts' comparison (similarity measurement) that factor can affect directly the scores and mislead the system. This leads to the necessity of keeping a copy of real text in a form similar to the one returned by the API. Hence, we remove all the punctuation and lowercase all letters in the real text. For the comparison we use the Jaccard similarity measurement, on a character-based level. If the Jaccard score of an interim result and a particular processed text phrase is over the threshold, the corresponding real phrase is shown.

For such systems it is always necessary to keep the computational cost low, while avoiding any extra delay on the process. It would be extremely costly and risky to compare each interim result with every single phrase of real text. For that reason, we introduce a moving window to assist our system minimize the risk and cost. The decision of the length of such a window is critical. A small window does not guarantee the natural flow in case of skipped phrases or system failure. On the other hand, the wider is the length of the window the more probable for the alignment system to fail. For that trivial definition we chose the length to be equal to five phrases per instance. The diagram in Figure 4 illustrates a simplistic form of our final system.

Figure 4
Live Captioning System



Source: Author's Illustration

While we iterate through the continuous flow of interim results produced, as long as the actor speaks, concurrently we calculate the Jaccard score with each element included in the window. The most obvious way to check if an interim result matches one of the elements is to calculate the similarity score for each element and then pick the element with the maximum similarity. If that score is beyond the threshold the corresponding phrase is tracked on the real unprocessed text and send on the screen. Based on that approach a significant problem may occur. In dialogues, it is common that the actors' lines share common phrases and that can lead in two or more phrases with high similarity scores. If we also consider the inevitable miss-transcribed phrases from the API, we could end with the common case shown in Table 1.

Table 1
Jaccard Similarity Scores of an Interim Result for Each Element of Processed Real Text

Jaccard Similarity Scores	**0.15**	**0.63**	**0.2**	**0.25**	**0.65**

Source: Authors Illustration

Here the maximum score is 0.65 and corresponds to the last element. If the threshold was 0.6, we would have confirmed the match and proceeded to the screen. However, with a threshold equal to 0.6 we can see that the second element also fulfils the matching criterion. In this case it is not clear which element is the right one. Therefore, to avoid such a scenario, one score is calculated and compared at a time with the threshold. If an element with higher score is found, we end the process and send it on the screen.

A challenging part of this research is to control the flow of the API responses. As we explained earlier the way that the recognizing service operates cannot identify the speaker diarization. Based on that, when the audio does not contain any gap of silence for more than a couple of seconds, the system cannot tag any result as final. Inevitably, we need to manipulate the API response at every stage of the recognition process. By identifying a result as final, the service recognizes that the upcoming audio has no connection with the previous audio so it stops stacking it as one phrase. When an interim result finally gets matched with one phrase in the real text, the number of words for both of them is calculated and the difference is held as information for the next comparison. The upcoming responses with lower or equal number of words should be ignored and not tested at all, whereas responses with greater number are firstly getting processed and then follow the testing procedure explained earlier. That process is the subtraction of the first n words. When the system marks a result as final, then this number is reset.

The way that the window moves across the real text while new API responses arrive and being matched is the following; when a match occurs, the window instantly changes with new starting point the matched element's position and range the size of 5 elements. It is worth noting that we have assumed the flow of theatrical plays does not have any backwards and if an actor gets confused the rectification point will be a next one. The constant tracking of the position in-between the real text is crucial in order to be able to transfer the information amongst the consecutive service requests.

Discussion

The value of the above research is not limited to the field of theatre but extends to other art forms, performing or not, that could benefit functionally and / or aesthetically from the introduction of such innovative elements. An equally important goal of the project is to investigate critical questions such as:

- From the creator's point of view, the artistic value of the proposed interventions. That is, whether these new tools can offer new aesthetic means of artistic creation.
- From the spectator's point of view, whether the introduction of such elements can make the theatrical experience more interesting, attractive and captivating. Children's theatre is of particular interest as it often includes highly imaginative and surreal elements while the children's audience seems more positively predisposed to such experiences.
- The role of these technologies in removing various barriers (accessibility, language, geography) and opening theatrical performances to important additional audience groups, and their contribution to increasing traffic.

For this purpose, in the framework of the project, two experimental theatrical performances will be held, which will be selected, adapted and augmented appropriately based on these new technological means. The first will be a children's play and will be an intermediate presentation of the results of the project, while the second will be an ancient tragedy and will incorporate a wider range of more developed related technologies.

Conclusion

The purpose of this research is to introduce new technologies and media in theatre and the performing arts, in order to offer an enhanced experience to the audience, but also to remove the obstacles that various groups face when accessing a theatrical performance, whether it is linguistic problems or sensory ones. For the above reasons we proceeded in the design and development of a system for monitoring the time flow of a performance. It is an advanced speech recognition system that monitors the progress of the performance by comparing the current dialogues of the actors with the script. By collecting the time stamps of the various cues, audio visual media, video projections and lighting, we can "wind them up" over the time flow so that they are performed automatically. By obeying in the same logic, various virtual characters will be included and will participate in the show, contributing aesthetic features beyond those that are possible by natural actors. For this purpose, two experimental performances will be held to decide on the robustness of the overall system, but also on whether these new tools can offer new aesthetic means of artistic creation. Finally, the most important goal of these technologies is to contribute to the removal of various barriers (accessibility, language, geography) and to be able to achieve the opening of theatrical performances to important additional audience groups.

References

1. Ars Electronica. (2006), "Le sacre du printemps" (The rite of spring), available at: https://ars.electronica.art/futurelab/en/project/le-sacre-du-printemps/ (May 22, 2020)
2. Bird, T. (2006), "Sunday in the Park with George", available at: http://www.projctn.com/sunday-park-george (22 April 2020)
3. Costea, G.-A. (2018), "Do you speak tech? Report from an interdisciplinary and cross-cultural partnership", available at: https://www.europeantheatrelab.eu/speak-tech-report-interdisciplinary-cross-cultural-partnership/ (22 May 2020)
4. Fiveash, K. (2016), "The Tempest review: Real-time digital avatar brews storm in a teacup", available at: https://arstechnica.com/gaming/2016/11/tempest-review-real-time-digital-avatar-performance-capture/ (22 April 2020)
5. Gao, J., Zhao, Q., Yan, Y. (2011), "Towards precise and robust automatic synchronization of live speech and its transcripts", Speech Communication, Vol. 53, No. 4, pp. 508-523.
6. 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, vol. 7, pp. 60968-60983.
7. Heimbach, H., Klein, A., Lin, A., Lu, M. (2016), "Developing theatrical lighting control with Arduino", available at: https://soe.rutgers.edu/sites/default/files/imce/pdfs/gset-2014/GSET2015-DevelopingTheatrical.pdf (22 May 2020)
8. Henry M. (2019), "Idiomatic", available at: https://theatredeliege.be/evenement/idiomatic/ (22 April 2020)
9. Mathewson, K. W., Mirowski, P. (2017), "Improvised theatre alongside artificial intelligences", The Thirteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment AIIDE-17, 5-9 October, Association for the Advancement of Artificial Intelligence, Palo Alto, California, pp. 66-72.
10. Perkowski, M., Sasao, T., Kim, J. H., Lukac, M., Allen, J., Gebauer, S. (2005), "Hahoe KAIST robot theatre: Learning rules of interactive robot behavior as a multiple-valued logic synthesis problem", 35th International Symposium on Multiple-Valued Logic ISMVL '05, 19-21 May, IEEE, Calgari, Canada, pp. 236-248.
11. Secară, A. (2018), "Surtitling and captioning for theatre and opera", In Pérez-González, L. (Ed.), The Routledge Handbook of Audiovisual Translation, Routledge, London, pp. 130-144.
12. Shaw D. (2012), "Transforming digital drama: the technology", available at: http://www.bbc.com/news/technology-17079364 (22 May 2020)
13. Siddiqui M. U. (2015), "Hakanai | Adrien M & Claire B", available at: https://www.arch2o.com/adrien-m-claire-b-hakanai/ (22 April 2020)
14. Thonon, J. (2018), "The re-invention of surtitling. a new technical approach to the international market", in Pfaud, T., Dusol, J., Wiley, H. (Eds.), Digital theatre - a casebook, European Theatre Convention, Berlin pp. 50-55.

About the authors

Charalampos Tsioustas received the Dipl.-Ing. Degree in Electrical and Computer Engineering from the National and Technical University of Athens, in 2018. He is working in the R&D Department of ITC SA and he is currently pursuing a second Master's Degree in Microsystems and Nanostructures. Author can be contacted at: **charis.tsioustas@itcnet.**

Daphne Petratou graduated in 2018 from the Department of Architectural Engineering of University of Patras and since 2018 she works in the R&D Department of ITC SA. She is competent in Computer Graphics, Architectural Design and Geographic Information Systems. Author can be contacted at email: **daphne@itcnet.gr.**

Maximos Kaliakatsos-Papakostas is a Research Associate at the Institute for Language and Speech Processing of Athena Research and Innovation Center (ILSP/ATHENA RIC), an Adjunct Lecturer at the Department of Technologies for the Digital Industries, National and Kapodistrian University of Athens and a Research Associate at the Department of Music Studies at the Aristotle University of Thessaloniki (AUTH). He has worked on research projects related to music and AI, (including FP7 COINVENT and a personal research grant from the Greek State Scholarships Foundation, among others). He has also worked on applied research project as a developer (H2020 iMuscica and NLP Theatre). He holds a PhD from the Department of Mathematics, University of Patras since 2014 where he graduated in 2006 and received his M. Sc. Diploma in Computational Intelligence since 2009. Author can be contacted at: **maximos@athenarc.gr.**

Vassilis Katsouros is currently Research Director and Director of the Institute for Language and Speech Processing of the Athena Research and Innovation Center, Athens, Greece. He received his Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 1992, the M.Sc. and Ph.D. degrees in electrical and electronics engineering from Imperial College, London, U.K., in 1993 and 1997, respectively. His research interests include signal processing, probability and statistics, image and audio processing, pattern recognition and machine learning. He is a Member of the IEEE, ACM and TEE. Author can be contacted at: **vsk@athenarc.gr.**

Apostolos Kastritsis received his diploma degree in Computer Engineering & Informatics from the University of Patras, Greece in 2008, received a diploma degree in Performing Arts & Music from the Indonesian Institute of Arts, Surakarta (Indonesia, 2014), achieved a M.Sc degree in Art, Virtual Reality and Multiuser Systems of Artistic Expression from Athens School of Fine Arts (Greece 2015) and Arts et Technologies de l'Image from the University Paris8 (France, 2016) . He also received a Bachelor's degree in Economics from the School of Economics & Business Administration (University of Patras, Greece, 2019). He is currently with the Institute for Language and Speech Processing and his research interests include human machine interaction, procedural animation, artificial intelligence, virtual and augmented reality and interactive systems. Author can be contacted at: **apostolos.kastritsis@athenarc.gr.**

Konstantinos Christantonis BSc, MSc, is a Research Assistant at the International Hellenic University. He works on projects involving intelligent systems. The main areas of his interest are natural language processing and data mining. He is a production and management engineer with an MSc in Data Science. Author can be contacted at: **christantonis.kons@gmail.com.**

Konstantinos Diamantaras received the Diploma degree from the National Technical University of Athens, Greece, in 1987 and the Ph.D. degree in Electrical Engineering from Princeton University, Princeton, NJ, USA, in 1992. From 1998 to 2005 he has been a faculty member at the Department of Information Technology, Technological Educational Institute of Thessaloniki, Greece. He is currently a Professor at the

Department of Information and Electronics Engineering, International Hellenic University, Greece. Author can be contacted at: **kdiamant@it.teithe.gr.**

Michael Loupis received a Dipl.-Ing. degree in Electrical Engineering from the Aristotle University of Thessaloniki, Greece, an M.Sc. degree in Microprocessor Engineering from the University of Bradford, U.K., a Ph.D. degree in Computer Engineering from the National Technical University of Athens, Greece, and an M.Sc. degree in Quality Assurance from the Greek Open University. Currently he is the Technical Manager of ITC S.A. and an Associate Professor with the National & Kapodistrian University of Athens. Author can be contacted at: **mloupis@uoa.gr.**