SMARTVOX – A WEB-BASED DISTRIBUTED MEDIA PLAYER AS NOTATION TOOL FOR CHORAL PRACTICES

Dr Jonathan Bell
Composer
Paris, France
belljonathan50@gmail.com

Benjamin Matuszewski
CICM/musidanse EA1572, Université Paris 8,
STMS Lab IRCAM-CNRS-UPMC
Paris, France
benjamin.matuszewski@ircam.fr

ABSTRACT

The present paper describes the features and implementation of *SmartVox*, [1] an application designed to help vocal ensembles learn and perform polyphonic music.

Technically, *SmartVox* is a distributed web application that delivers audiovisual scores through the performer's mobile devices. From a singer's point of view, this setup allows for the synergy between visual and acoustic stimuli, which facilitates the interpretive and performative processes, particularly in polyphonic passages. It also enables spatial separation of the performers (*cori spezzati*), and speeds up the learning process of unfamiliar musical materials (e.g. microtonal tuning, texts in a foreign language).

The ubiquity of smartphones makes such a distributed system affordable and allows the use of *SmartVox* in multiple contexts, from professional ensembles to pedagogical and recreational practices.

Introduction

Related Work

Many composers in the twentieth century have designed audio systems in order to conduct performers. The first extensive use of technology-assisted conducting dates back to the 1920s, when click tracks were used in order to sync the orchestra to silent movies. [2] In the domain of experimental music, the first and most explicit research is found in Emmanuel Ghent's *Programmed Signals to Performers, a New Compositional Resource* [1], where the author proposed a system derived from the click track technique. This system, the *Coordinome*, was able to send audio signals pre-recorded on a magnetic tape individually to each performer.

During the second half of the century, many systems were developed to guide performers with multiple click tracks in order to realize complex polytempo pieces (Elliot Carter's *2nd String Quartet*, Brian Ferneyhough's *Mort Subite*, Karlheinz Stockhausen's *Helicopter Quartet*, Iannis Xenakis' *Persephassa* to name just a few).

More recently, Phillipe Kocher developed a technology-assisted conducting system for the realisation of polytempo networks [2]. Similarly, the Latvian composer Rytis Mazulis used in many pieces equivalent systems, sending to performers extremely slow glissandi generated electronically [3].

Figure 1. *Au Commencement* (J. Bell, 2016), performed by the *Mangata* ensemble using *SmartVox*.

Performance-Centrism

In most of the above-mentioned cases, technologies of assisted conducting were created to solve *composition-centered* issues: concerned with the realisation of a specific musical material. This focus on composition, to the detriment of its *performative* counterpart, draws attention to difficulties that can be encountered by the person who plays/sings the music. Indeed, the use of click tracks (audio) or animated notation (visual) changes the traditional relationship established between composer, conductor, and performer. In a musical ensemble, these forms of extended notation often imply the replacement of a human conductor by a technological solution. The improvements in precision that accompany the use of such systems do not come without drawbacks. For example, polytempo and microtonal passages become simple to realize, even with musicians placed around the audience, but the stimulation of the player by a live conductor is not guaranteed any more. The challenge for the composer therefore consists of find-

[1] https://github.com/belljonathan50/SmartVox0.1 – the term *SmartVox* was coined by Laurence Brisset from the *De Caelis* Ensemble.

[2] The invention of the click track is usually credited to Max Steiner, although other sources have attributed it to Scott Bradley or Carl Stalling.

ing ways for these new forms of notation to remain as efficient as more conventional setups. According to this *performance-centric* point of view, the singer's reception, appropriation and feedback play an essential role in the assessment of the application.

1. AUDIO-SCORES AND ANIMATED NOTATION

SmartVox is characterised by the use of audiovisual (i.e. multimedia) notation. This recent development was initiated by preliminary research on *audio-scores* [4] and *animated notation* (screen-scores). [3]

1.1 Audio-Scores

Microtonality and polytempo are two realms of investigation which often lead composers to extend their notation with an acoustic element (such as click track, sine-tones, or any other auditory information) [1][2][3][4]. These signals are generally sent through an earpiece, so that they only concern the performer and not the audience. Such audio cues can be particularly helpful for vocalists. Indeed atonality and microtonality impose even more challenges to singers than to instrumentalists, because singers cannot rely on gestural automatisms to find their pitch and adjust their intonation to others. Writing microtonal music for *a cappella* voices, or placing musicians in the performance space therefore present difficulties that audio-scores attempt to solve, using audio technologies as a means to 'augment' the traditional notation. This compositional research was initially motivated by the intuition that the coupling of auditory and notational media would give the performers access to new and expressive situations, thus investigating the fields of computer-aided composition and computer-aided performance. [4]

In comparison to traditional sheet music, an audio-guide is often initially surprising for the singer, but quickly becomes a very useful cue which one can rely on. The performer is not only asked to reconstruct his part from the symbols he reads, but also to imitate what is heard through the earpiece. The notational input is therefore not only visual, but also auditory, hence the term *audio-score*. In previous work with the *De Caelis* ensemble over a period of ten years, [5] audio-scores have proved to convey an invaluable tool for the learning process of a piece of microtonal music. Audio-scores can also be very effective in performance since they simplify the task of the singers, free them from tuning forks and from the anxiety of getting lost in a difficult passage, both in terms of intonation or temporal coordination.

In this work however, audio-scores were always thought of as an extension of, rather than an alternative to notation. One of the greatest assets of musing-reading is that it allows the performer to preempt upcoming difficulties, whereas audio fluxes do not.

[3] http://animatednotation.com/composers.html

[4] The term of 'Computer-Aided Performance' was coined by M. Malt for the realisation of John Cage's *Concert for Piano and Orchestra* [5].

[5] The *De Caelis* ensemble has actively supported the project and commissioned five pieces using such systems.

1.2 Screen-Scores

In both experimental music and classical repertoire, tablets are increasingly replacing sheet music. However, the size of phones or tablet screens rarely matches that of large printed scores, and screen-based notation typically requires large font sizes and more page turns than its printed counterpart. A balance must therefore be found between displaying what is executed in the moment, and what is coming next, so as to always convey the most useful information to the performer. Indeed, problems may arise if a difficult passage (e.g. a high pitch or a long phrase) suddenly pops up at the start (at the left-end) of the page: breathing, or preparing an attack on a certain pitch, in sight-reading situations particularly, can require a few seconds of preparation that must be taken into account when creating the animated score.

In screen-based animated notation, the representation of time often contrasts with traditional sheet music. *SmartVox* for instance, uses notation realised in *bach.roll*, a Max object for proportional notation, with a cursor moving from left to right. This unfolding of time, inspired by digital audio workstations, offers an intuitive representation of speed and duration of musical events. Although it differs drastically from the classical "bars and beats" notation of rhythm, this solution proved to be very useful for singers since they need to anticipate the duration of a musical phrase to come in order to control their breath accordingly.

This proportional representation of time also profoundly simplifies temporal coordination between performers, particularly in non-pulsed music. For an inexperienced singer for instance, the main advantage of having his part displayed on a dedicated screen is the fact that he cannot get lost. Indeed, the screen score typically displays at any given moment only what is happening in his own part.

Such forms of augmented notation are therefore very convenient when singers have long parts to learn, when they are distanced from one another in the performance space (e.g. around the audience), or if the piece requires singing with microtonal intervals.

1.3 Early Prototypes

The *SmartVox* system was preceded by a series of prototypes with similar functionalities, developed over the past years.

Wired Cable Systems

The piece *Deserts* (2007) [6] used a computer to send audio-scores realized on a digital audio workstation. In this setup, each performer received two types of audio cues through an earpiece:

- A sequence of microtonal pitches.
- A series of clicks in an individual tempo.

Several techniques were investigated to distribute the individual audio cues to the performers, including mp3 players and wireless systems such as infra-red and high-frequency headphones as well as Bluetooth-based systems.

[6] Composed by J.Bell and commissioned by the *De Caelis* Ensemble

However, these systems were not robust enough to be deployed in concert performances so that wired systems, that usually consisted of a multichannel sound-card wired to headphones, were always preferred in such situations. These systems therefore locked the number of performers to the available output channels and impeded the ability to distribute the performers in the concert area.

Wireless Native Application

The piece *De Joye Interdict* (2014), also commissioned by the *De Caelis* Ensemble, used *iVideoShow*, a commercially available iOS application. The wireless setup was robust enough to guide performers in a concert situation. The application allowed the playback of individual video sequences on the iOS devices remotely controlled through Open Sound Control messages sent over a Wi-Fi network. This allowed the performers to be placed at a significant distance from each other. However, however *iVideoShow* presented several limitations:

- The system was difficult to improve and to deploy over cross-platform devices (iOS only).
- The synthesized audio cues, realized on a digital audio workstation, were difficult to control.
- The graphical notation captured as static screenshots from a notation software remained very close to traditional sheet music.

2. TECHNICAL SETUP / DESCRIPTION

The *SmartVox* application is based on the *Soundworks* framework [6]. The audiovisual scores – distributed as simple video files – are produced using the *Bach* (Bach Automated Composer's Helper) [7] environment in Max. [7]

2.1 Generating Audiovisual Scores Using the *Bach* Environment

The audiovisual scores used in the application are realized with *Bach*, a Max library for real-time computer-aided composition. In this environment (*bach.roll* or *bach.score*), each note of the score can be associated with metadata. Here, the feature is used to configure and control the synthesizer [8] that creates the audio cues, directly from the notational environment.

During the composition process, this particular setup allows for a workflow that consists in sculpting each vocal line with its appropriate pitches (in eighth-tones), text, intensity curves, formants, elocution velocity, glissandi, with a real-time audio feedback (see Fig. 2, 3).

The audio-score often only sounds when the performer should sing, yet some useful information can be provided to the performers during long silences: these audiovisual cues can deliver in advance the musical phrase that is coming next. In Figure 4, the lower stave is an anticipatory cue (the performer just listens), and the upper stave is an audio-guide (the performer sings along with the earpiece).

[7] https://cycling74.com/

[8] The *psych* module for Max - this module performs high quality pitch correction (auto-tune) and polyphonic harmonizing of monophonic audio sources.

Figure 2. A sample of spoken text stored inside the *filename* slot of each note (or group of notes) of the *bach.roll* object.

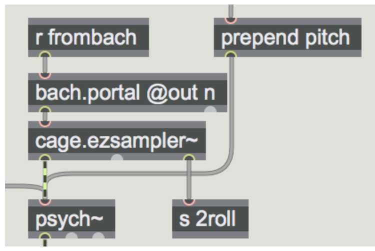

Figure 3. Excerpt from a *Max* patch showing the *cage.ezsampler* object retrieving metadata stored in the *Bach* score. The extracted information is then sent to the *psych* object, which transposes the given sample of spoken text to the defined microtonal pitch.

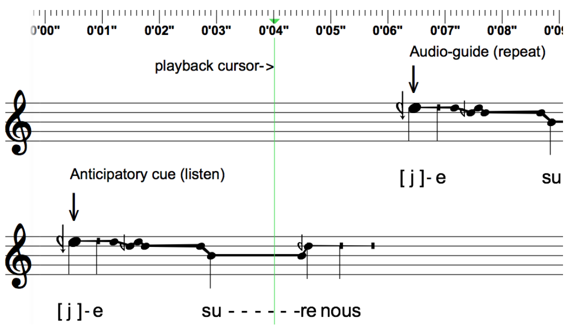

Figure 4. Screenshot of the video captured from the rendering of the *Bach* environment and played back by the mobile application based on the *Soundworks* framework.

2.2 Distributing Individual Scores to the Performers Using the *Soundworks* Framework

The system that distributes the audiovisual scores created with *Bach* is entirely based on web technologies, and more specifically on the *Soundworks* framework. [9] *Soundworks* provides a set of services – such as synchronization, network messages, distributed states, creation of groups of clients – that aims to solve problems common to distributed and synchronized web applications centered on multimedia rendering. The framework is written in *Javascript*, with a server side based on *NodeJS*. [10]

The *SmartVox* application consists of two web clients, the *player* and the *conductor*, that can be executed in any recent web browser on mobile devices (e.g. smartphones, tablets) and laptops. The real-time communication between

[9] *Soundworks* has been developed in the framework of the CoSiMa research project funded by the French National Research Agency (ANR) and coordinated by Ircam.

[10] https://nodejs.org/en

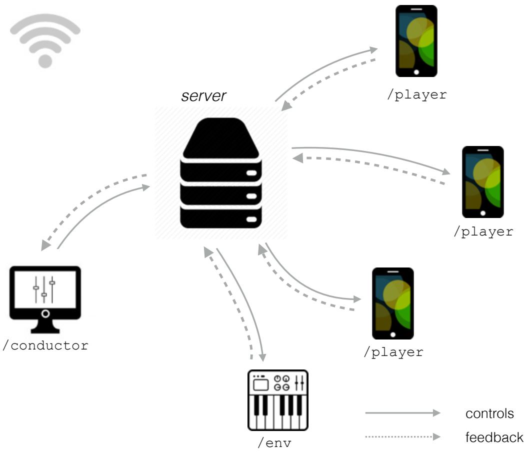

Figure 5. Architecture of the *SmartVox* application.

clients is achieved through the *WebSocket* protocol [11] (see Fig. 5).

*The **player** client*

This client, dedicated to the performers, is essentially a remotely controlled and synchronized video player.

Figure 6. Screenshot of the *player* client - drop-down menu showing the available parts of the score.

When entering the application, the performers are required to choose their part among the available ones (see Fig. 6). Once done, the corresponding audiovisual score is sent to the performers by the server. When the video is received on the device, further interactions with the score are locked in order to prevent the performers from accidentally changing the temporal position in the video, and thus ensure correct temporal coordination among all performers.

Additionally, this client can be used for the rendering of audio files (mp3) and/or videos (mp4) dedicated to the audience, through loud speakers and projectors.

*The **conductor** client*

The second client is dedicated to the choirmaster (see Fig. 7). Its role is to control the global and distributed state of the application. Through this interface, the conductor can therefore control the playback of the audiovisual scores:

- Start, pause and stop the video.

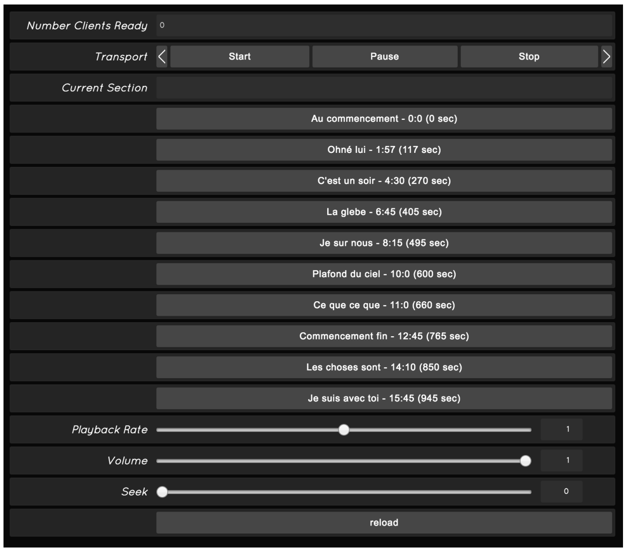

Figure 7. Screenshot of the *conductor* client.

- Jump to a labeled section of the piece or to a specific playback time.
- Change the playback rate (i.e., the speed) of the video without altering the pitch.
- Change the volume of all connected clients.

The interface also provides feedback (e.g., number of connected clients) to the choirmaster - information that proved to be of primary importance in concert situations.

Configuration

The application can be configured through a data structure that defines the path to the video files as well as the positions and labels of the sections of the piece. This allows for easily adapting the application to the content of a given piece:

```
const score = {
  duration: 20 * 60, // seconds

  // define the different parts
  parts: {
    'soprano-1': {
      file: 'videos/soprano-1.mp4',
    },
    'soprano-2': {
      file: 'videos/soprano-2.mp4',
    },
    // ...
  },

  // define the different sections
  sections: {
    alpha: {
      time: 0,
      label: 'First␣section',
    },
    beta: {
      time: 117,
      label: 'Second␣section',
    },
    // ...
  },
};
```

Both interfaces are dynamically generated according to this configuration, the list of sections in the *conductor* interface, and the list of available parts in the *player* interface.

[11] https://www.w3.org/TR/websockets/

2.3 Case Report: *Au commencement*

The application has been prototyped in parallel with the composition and rehearsal of the piece *Au Commencement* by the *Mangata* ensemble. One of the objectives of the project was to help the singers to tune-in (*i.e.* match spectrally) with the fixed-media electronics. The wireless and cross-platform system allowed singers to do so, while being placed around the audience (see Fig. 8).

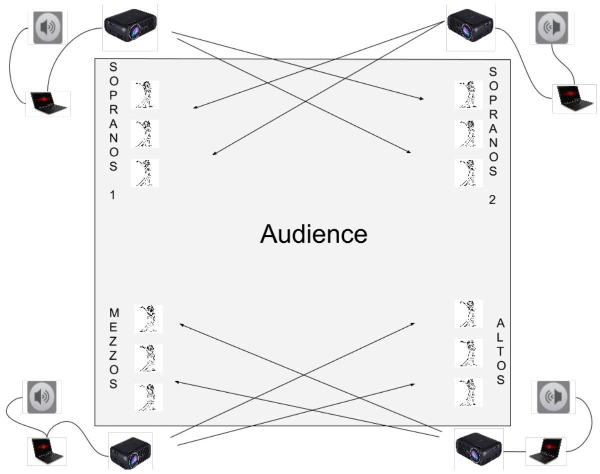

Figure 8. Spatial configuration of performers, lightings and loudspeakers.

A first public rehearsal of the piece revealed several issues with the user interface design. Indeed, the setup was already stable in rehearsals, but the challenges that arose in the public performance still required several modifications. Discussions with the choirmaster and the performers revealed the source of these problems: the most significant feedback concerned the lack of information about the current state of the system (in both *player*'s and *conductor*'s interfaces). As a result, several features were introduced in the next version of the application:

- On the *conductor* interface, display of the number of connected players, and addition of a button allowing to reinitialize all connected players.
- On the *player* interface, display of the name of the chosen part, and removal of the video controls.

The premiere of the piece took place in the *Notre Dame de Bon Secours* church in Paris [12] . Four groups of singers were situated at a significant distance from each other. Each of the twelve singers had a smartphone and earphones. Four additional computers, acting as generic *player* clients and connected to loudspeakers and projectors were placed in the four corners of the area (see Fig. 8). The technical setup confirmed the convenience and cheap cost of wireless cross-platform web technology: in spite of the disparity of the users' devices, communication was established quickly through the web browser, between Android phones, iPads, iPhones (the singer's devices), OSX and Windows (environmental videos).

Acoustically, from an audience point of view, the extreme spatial separation between sources produced an immersive feeling and clarified the listening experience. This panoramic sonic image nevertheless let voices and electronics blend together successfully, thanks to the harmonic/spectral match between the two, and because of the natural reverberation of the church.

3. DISCUSSION

3.1 Pros

Technically, in its current state, the application demonstrates promising assets:

- Using web standards and Node.js, the application can be executed on a large range of platforms and devices.
- The architecture of the application allows for easily adapting it to different pieces by modifying its configuration.
- The application can be setup effortlessly and quickly in rehearsals and concerts using a laptop running the server and a Wi-Fi router.
- During rehearsals, the conductor client allows for flexibly navigating within a given piece.

From the performer's point of view, *SmartVox* is perceived as a useful device which combines acoustic and visual stimuli to help interpret challenging polyphonic scores, wether in rehearsals, performance, or pedagogical contexts. Also, experience proved that audiovisual scores surprise groups of performers in a positive way, which can be used as an impulse for challenging and imaginative musical/performative experiments.

Finally, from a compositional point of view, the piece *Au Commencement* demonstrated that the setup accelerates the learning process, facilitates the realisation of microtones, and allows to place performers at a large distance from one another, without putting them at risk in performance.

3.2 Cons

Feedback from the singers showed that audio cues sent through earpieces, while being useful in difficult passages, can tend to impede mutual listening. Further experiments will therefore propose visual cues only (or audio cues only), depending on the musical passage, in order to optimize the quality of the information given to the performer. [13]

Also, the clock synchronization between clients provided by the *Soundworks* framework [8] should be integrated in the application. Such precise time control among clients would thus allow for more complex rhythmical writing.

3.3 Use Cases

SmartVox owes its development to professional ensembles (*De Caelis* and *Mangata*) who tested the prototype in rehearsals and concerts, and gave invaluable feedback at each

[12] A recording of the premiere of the piece is available at: https://youtu.be/uVGPa1Z6Ji8

[13] The exact delimitations of these cues could be defined iteratively from discussions with the singers and choirmaster together with an A/B testing strategy to speed up the process.

stage of its development. Since then, however, the application has also been used in different contexts:

Pedagogy

Recent experiments in conservatoires [14] demonstrated that, for children, a distributed mobile application can be evocative of a multiplayer game. This playful aspect helps to focus their attention on challenging music theory notions (*solfege*). In a pedagogical piece composed for this system, the notation purposefully conveyed the same pitch information in four different ways:

- Sound frequency: a synthetic voice sings on a given pitch, e.g., A = 440 Hz (audio).
- Spoken words: the synthetic voice pronounces the corresponding phoneme '*La*' (audio).
- Symbolic notation: the corresponding pitch is showed on the musical stave (visual).
- Written text: the phoneme '*La*' is written below the stave (visual).

After a few sessions, groups of ten to twenty children were able to sing in tune complex three-parts polyphonies.

For older students (undergraduates), *SmartVox* has been used to sing extracts of a motets of the Ars Nova, psalms of the Renaissance, as well as examples of early *solmisation*.

Amateur Choirs

A cappella choral singing requires competencies such as vocal skills, intonation and music reading. This often restricts ancient and contemporary repertoire to a small group of specialists. The audio and visual guides provided by *SmartVox* therefore seek to give accessibility and exposure to works otherwise judged too difficult (e.g., motets of the Renaissance or atonal music), for choirs of all levels.

Conclusion and Future Work

The present article described *SmartVox*, a web-based application specifically designed to help choristers in the realization of challenging pieces (i.e. including intricate performative aspects such as microtonality and unusual placement of performers in space). The application proved to be succesful in rehearsals, performances, as well as in pedagogical contexts.

Currently, the application is deployed over a private wireless local area network through Wi-Fi. The connected clients therefore require the physical presence of the server where the performance takes place. However, the application being completely implemented based on standard web technologies, it could be easily hosted on a public remote server and thus accessed over the Internet. While not suitable in concert situations, this feature will make an important difference in terms of spontaneous access and dissemination of the application, especially in pedagogical contexts.

In future versions, the application could also be expanded to allow for a better appropriation by the performer. The audiovisual score could then be used by the choirmaster during the rehearsal process, to make notations about the playback of the video (e.g. *accel.*, *fermata*, *crescendo...*). This would then be communicated to the performers.

The positive results thus far encourage us to believe that *SmartVox* will continue to be an innovative and useful form of musical notation.

[14] The recording of a *SmartKids* reading session is available at: https://youtu.be/hlHAeiWT28Y

Acknowledgments

We are deeply grateful to the following people for their precious support and advice: Andrea Agostini, Julian Anderson, Richard Baker, Eric Daubresse, Stefano Gervasoni, Daniele Ghisi, Marisa Gupta, Bryn Harrison, Gregoire Lorieux, Paul Newland, Hector Parra, Nye Perry, Norbert Schnell, and James Weeks.

References

[1] E. Ghent, "Programmed Signals to Performers: A New Compositional Resource," *Perspectives of New Music*, vol. 6, no. 1, pp. 96–106, 1967.

[2] P. Kocher, "Polytempo Network: A System for Technology-Assisted Conducting," Institute for Computer Music and Sound Technology, Zurich University of the Arts, 2014.

[3] G. Daunoraviciene, "Sound architecture of Rytis Mažulis' microstructural canons (from 100 to the 3,448275862 cents)," *Menotyra*, vol. 30, no. 1, 2003.

[4] J. Bell, "Audio-Scores, a Resource for Composition and Computer-Aided Performance," Composition Doctorate, Guildhall School of Music and Drama, 2016.

[5] B. Sluchin and M. Malt, "Interpretation and Computer Assistance in John Cage's Concert for Piano and Orchestra (1957-1958)," in *7th Sound and Music Computing Conference Proceedings*, 2010.

[6] S. Robaszkiewicz and N. Schnell, "Soundworks – A playground for artists and developers to create collaborative mobile web performances," *http://architexte.ircam.fr/*, 2015.

[7] A. Agostini and D. Ghisi, "Real-time Computer-Aided Composition with BACH," *Contemporary Music Review*, vol. 32, no. 1, April 2013.

[8] J. P. Lambert, S. Robaszkiewicz, and N. Schnell, "Synchronisation for Distributed Audio Rendering over Heterogeneous Devices, in HTML5," in *Proceedings of the 2nd Web Audio Conference (WAC-2016)*, 2016, pp. 6–11.