KEER2010, PARIS | MARCH 2-4 2010

INTERNATIONAL CONFERENCE ON KANSEI ENGINEERING AND EMOTION RESEARCH 2010

ACTIVE MUSIC EXPERIENCE USING MOBILE PHONES

Maurizio Mancini [*] [a], **Antonio Camurri**[a], **Giovanna Varni**[a] **and Gualtiero Volpe**[a]

[a]*InfoMus Lab, DIST, Università di Genova, Italy*

ABSTRACT

This paper describes novel forms of social interaction based on music and gesture, with a special focus on KANSEI aspects such as real-time analysis and processing of non-verbal expressive, emotional content conveyed by human full-body movement and gesture. We present two scenarios toward mobile applications for active experience of sound and music content: Mobile Orchestra Explorer and Sync'n'Move. We also present a questionnaire-based evaluation of these scenarios that resulted in overall user positive feedback and will be useful for improving our interaction paradigm.

Keywords: KANSEI Information Processing, active music listening, HCI

1. INTRODUCTION

One of the challenges for novel interaction paradigms in Human Computer Interaction is the design and creation of a new type of systems, able to detect the user emotional state and to establish an *Affect Sensitive* interaction with the user, in the sense defined by, for example, Affective Computing [1] and KANSEI Information processing [2]. KANSEI Information Processing refers to coding and decoding of feelings, intuition and sympathy in Human-Computer Interaction in a larger perspective compared to Affective Computing: while AC is more concerned with emotions, KANSEI rather refers to a wide collection emotion related aspects, e.g., moods, feelings and so on.

This paper describes two music active listening scenarios using mobile phones. Active listening is the basic concept for developing interactive music systems which are particularly addressed to a public of beginners, naïve, and inexperienced users. User non-verbal behavior and gesture are measured to modify in real-time music content: for example, mobile phones are used to detect

[*]**Corresponding author:** Maurizio Mancini, InfoMus Lab, DIST, Università di Genova, Italy, Viale Causa, 13, I-16145 - Genova (Italy), maurizio.mancini@dist.unige.it

the movement of the user, to activate and control the music sections; or users move rhythmically and the phase synchronization extracted from their gesture drives the orchestration and rendering of a pre-recorded music. At the KANSEI level, user gestures are analyzed in order to extract their expressive, emotional content. Such information is used to modify in real-time the emotional content of music, e.g., by exploring different expressive performances of the same music piece (e.g., an intimate, an extrovert, a solemn, and an angry one).

In the first scenario we present in this paper, called Mobile Orchestra Explorer, the user moves in a virtual space occupied by an orchestra: as he gets close to an instrumental section, he gradually listens to such section. In the second scenario, called Sync'n'Move, two users interact by freely moving their mobile phones while listening to a piece of music: every time the users are successful in being synchronized the music orchestration and rendering is enhanced.

In Section 2 and 3 we provide a detailed descriptions of these two scenarios and in Section 4 we present an evaluation study of the above scenarios based on data collected with questionnaires filled up by participants attending the Agora Festival 2009 (IRCAM, Paris, June 2009).

2. MOBILE ORCHESTRA EXPLORER

2.1. Synopsis

The Mobile Orchestra Explorer paradigm enables active experience of prerecorded music: the users can navigate and express themselves in a shared (physical or virtual) "orchestra space", populated by the sections or single instruments of a prerecorded music: a user can activate and listen to one or more sections of the music. The mobile phones are here used to detect the movement of the user, to activate and control the music sections, and to present, on the phone display, user's position in the orchestra space. The music rendering is either based on 2D sound via loudspeakers (using WFS) or on the mobile phone using its headphones.

2.2. Technical description

The user stands in a room equipped with a 2D spatial sound speaker, holding his mobile phone in his hand. He can see on the display of the mobile a simple graphical representation of the virtual orchestra, with icons showing the position of instruments/sections in the space. A red cursor on the mobile display follows the user position in the orchestra space. The user task consists to explore this space to discover the music that the various instruments are playing. By tilting the phone left-right or forward-backward, the user moves the red cursor horizontally and vertically in the orchestra space: when the user moves the red cursor closer to instrumental sections on the display, he gradually listen to such sections. When the user performs expressive gestures, music expressive content is molded accordingly. For example, a harsh, abrupt gesture causes the instrumental section to be "frightened" and a deformation of the sound is heard. Recognition of different expressive contents causes exploration of different expressive performances.

The resulting sound is rendered in 2D, that is, the orchestra space is mapped on the physical space the user occupies in the room. Alternatively, the user can experience also using headphones.

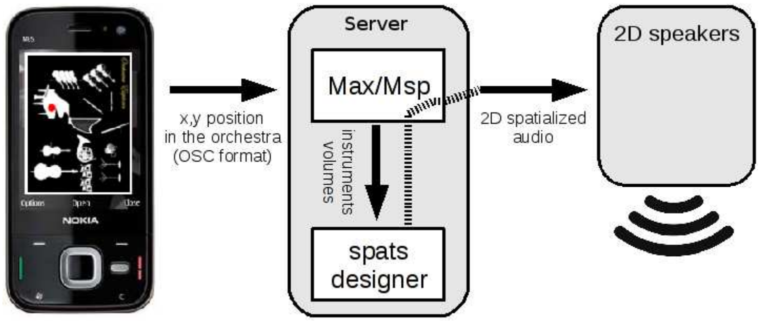

Figure 1: Mobile Orchestra Explorer architecture when using the 2D spatializer speaker: the user moves in the orchestra space by tilting the mobile phone; user position is sent to Max/MSP for audio generation and audio is produced by the 2D spatializer speaker.

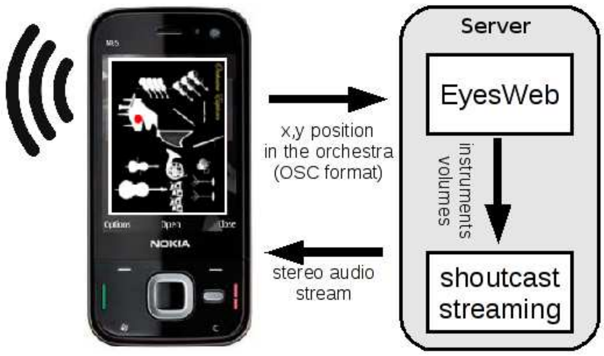

Figure 2: Mobile Orchestra Explorer architecture when audio rendering is sent back to the mobile phone: user moves in the orchestra space by tilting the mobile phone; user position is sent to Max/MSP for audio generation and audio is sent back in real-time to the user mobile phone.

In this case, the music sections are rendered directly by the mobile phone. The Orchestra Explorer architecture is shown in the Figure 1.

A module on the mobile phone written in Phyton computes the user position in the Orchestra space by integrating over the acceleration on the X and Y axis measured by the accelerometer in the phone. It also draws the Orchestra diagram and updates the red dot representing the user position. Then the dot coordinates (x, y) are sent to Max/MSP [3] through an UDP packet containing an OSC message containing two integer numbers ($[0, 240]$ for the Y axis, $[0, 320]$ for the X axis).

The spats designer plugin in Max/MSP, depending on the user position in the Orchestra space, determines which instruments can hear by the user and with which volumes. Finally Max sends the audio to a speaker able to render the instruments sound in 2D. If the 2D speaker is not available, the system is modified as shown in Figure 2.

In this case the (x, y) coordinates in the Orchestra space are sent to an EyesWeb XMI [4] application which produces the stereo rendering of the instruments near the user position. The resulting audio is broadcasted on the network through a Shoutcast server implemented in EyesWeb XMI and is received on the mobile phone where the user can hear it with headphones.

Figure 3: (upper) the example of the map representing the "orchestra space" on the mobile phone display; (lower left) a user is navigating in the Orchestra; (lower right) a user is exploring the "orchestra space" by moving the red cursor, to experience active listening based on the Orchestra Explorer paradigm.

An video example of how the scenario works is reported at the address:

`www.sameproject.eu/Demo`

Figure 3 shows user interacting with the Mobile Orchestra Explorer scenario during the public event Agora Festival 2009.

3. SYNC'N'MOVE

3.1. Synopsis

Sync'n'Move enables users to experience novel form of social interaction based on music and gesture, using mobile phones as a social interface [5]. Users need only to move rhythmically their mobile phones, for example, by holding them in their hands or by wearing them (e.g., putting the mobile in a pocket) and dancing. The users phase synchronization degree, extracted from their gestures, is measured and used to modify in real-time the performance of a pre-recorded music. This is a first example of shared collaborative active music listening experience. When the users synchronize successfully, the music rendering is enhanced, that is, new sections start to play; while in case of low synchronization, i.e., poor collaborative interaction, the music gradually looses sections and rendering features, until it becomes a very poor monophonic audio signal.

3.2. Technical description

Figure 4 sketches out how Sync'n'Move scenario works.

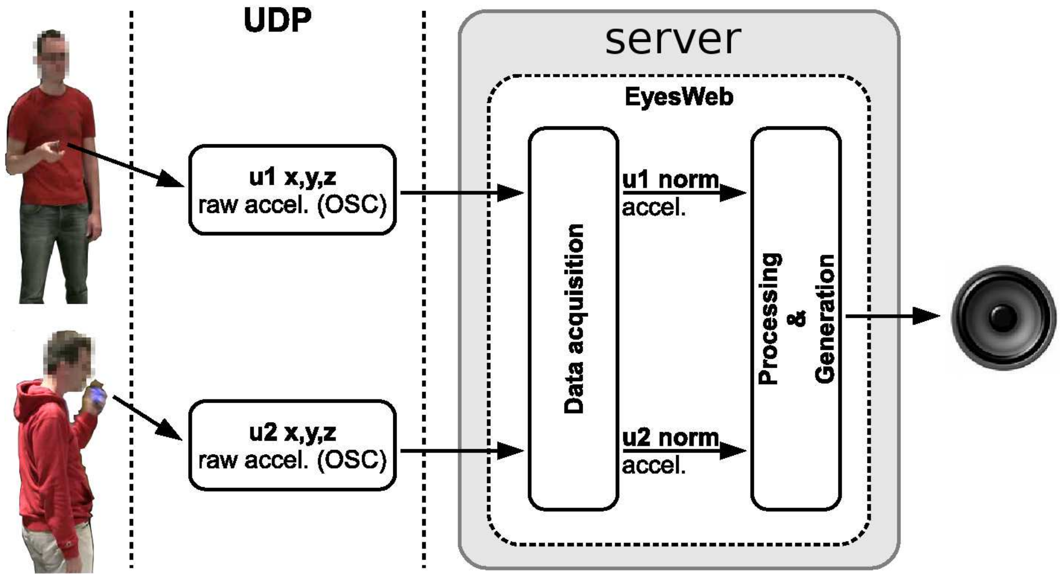

Figure 4: Sync'n'Move architecture: users move their mobiles, the raw accelerations are detected and analyzed. A multitrack audio content is generated according to the phase synchronization degree reached by the users.

Two users freely move their mobile phones and their hand/body acceleration is detected and measured by the tri-axial accelerometer embedded in the mobile. As one or both users begin to move or shake their phones they hear a drum sound playing a continuous rhythmic pattern. The users task is to perform movements with the phones in space trying to synchronize on the same rhythm, e.g., the one performed by the drum. During the task, an index of phase synchronization is extracted in real time from the users'gesture: every time the users are successful in being synchronized (index is high) then a bass, a guitar and a voice start to play together with the drum. If they do not succeed or they interrupt their interaction (index low) synchronization then only the drum continues playing.

Each mobile runs a Python script that collects data from the accelerometer and creates an OSC packet. This packet is sent to an EyesWeb XMI application composed by the following two main modules.

The *Data Acquisition* module acquires, calibrates and computes the normalized acceleration sent by the mobile phones the users are moving.

The *Processing & Generation* module computes the phase synchronization index CPR (Correlation Probability of Recurrence) exploiting the concept of Recurrence [6] and Recurrence Quantification Analysis [7], [8]. First of all, recurrences are detected in the accelerations using a small cutoff distance ε. Secondly, the probabilities that the accelerations recur after τ time steps at a certain value are computed. Then, the module computes the CPR index, defined as [9]:

$$CPR = \langle \bar{p}_{\vec{x}}(\varepsilon, \tau) \bar{p}_{\vec{y}}(\varepsilon, \tau) \rangle \tag{1}$$

where $\bar{p}_{\vec{x}}(\varepsilon, \tau)$ and $\bar{p}_{\vec{y}}(\varepsilon, \tau)$ are the probabilities of recurrence of acceleration normalized to zero mean and unitary standard deviation.

Finally audio content is produced: the audio content changes according to the synchronization degree between the users and the following three cases can occur:

- no audio: the users are not interacting at all, that is they are not moving their mobile phones. In this case the Processing & Generation block detects that the two accelerations are equal to zero and inhibit audio generation;
- metronomic audio: (i) only one user is moving or (ii) both are moving but in a not synchronized way. In the first case, the Processing & Generation block detects that just one of the accelerations is not from zero and enables the generation of a metronomic section in the audio output, e.g. drum instrument. In the second one, both the accelerations are analyzed and the CPR is computed. However, it is too low to allow the generation of the full audio output.
- full audio: the two users are moving in a synchronized way. The CPR assumes a high value (almost one) and the time for which the two users maintain this synchronization regime is measured. According to the length of this time, new sections are increasingly added to the audio content: the longer is the synchronization time the larger is the number of the enabled instruments, e.g. drums, bass and guitar, voice.

An video example of how the scenario works is reported at the address:

`www.sameproject.eu/Demo`

Figure 5 shows user interacting with the Sync'n'Move scenario during the public event Agora Festival 2009.

4. EVALUATION

An evaluation test of the Mobile Orchestra Explorer and the Sync'n'Move scenarios was carried out via anonymous questionnaire in the framework of the multidisciplinary encounter Agora Festival 2009 having as participants the event attendees. The questionnaire was designed by the SAME Consortium and it both collected information about users such as age, gender, nationality, work and addressed questions about the scenarios. More specifically, participants were asked to express their ratings on understanding, control, interaction level, fun, interest, future exploitation, engagement, and pleasure scales divided into eleven steps ranging from *not at all* to *very*. A blank

Figure 5: (left) two users try to synchronize performing the same kind of gesture (mirror); (right) two users synchronize by performing different gestures but with the same rhythm.

space was left in the last page of the questionnaire for comments and suggestions. The participants were classified in two groups: expert and non-expert users. Expert users were people having attested expertise in the music field, such as composers, music teachers, and professional musicians. All the participants applied voluntarily to this task and they were only asked to have a spontaneous behavior as much as possible. Each of them tried the scenarios for the first time and there was not any previous acquaintance with the scenarios. Before their performance, they were provided with a short demonstration of how the scenarios work. More details on participants will be provided in the next subsections. We took into account partially filled up questionnaires also.

4.1. Mobile Orchestra Explorer evaluation

4.1.1. Participants

A group of 26 people (22 male and 4 female from different european and extra-european countries) evaluated this scenario. Mean age of the participants was 34.8 years (from 18y to 64y). The expert users group was composed by 12 people.

4.1.2. Results

Figure 6 summarizes via box plot the results for the Mobile Orchestra Explorer scenario. From the inspection of this plot we can infer that globally the scenario was very easy to understand (median=10). Figure 7 depicts the two box plots for evaluation by non-expert (left) and expert users (right), respectively. The y-axis shows the ratings range expressed as numerical values (0 for *not at all*, 10 for *very*),whereas the x-axis shows all the items from the questionnaire *understanding*, *control*, *interaction*, *fun*, *interest*, *future*, *engagement*, *pleasure*. Non-experts gave significantly higher and less spread ratings than those provided by experts. More specifically, non-expert users reported they had greater control over the scenario. Further, interaction and engagement showed a very high difference between the ratings from the two groups: the middle 50% of the ranked ratings for interaction ranges from 6 to 9 for non-expert users and from 2 to 8 for experts, whereas the middle 50% of the ranked ratings for *engagement* ranges from 6 to 9 for non-expert users and from 2 to 8 for experts. This may be due to the fact that the Mobile Orchestra Explorer also has an educational valence, i.e., it explains how an orchestra plays: which are the sections, how they interact, how the score and the single instruments are organized in time to build the music

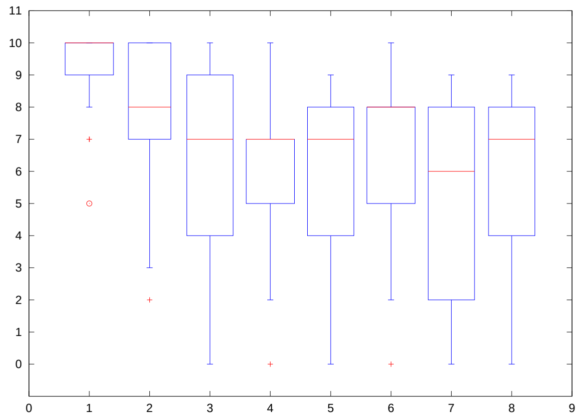

Figure 6: The global box plot

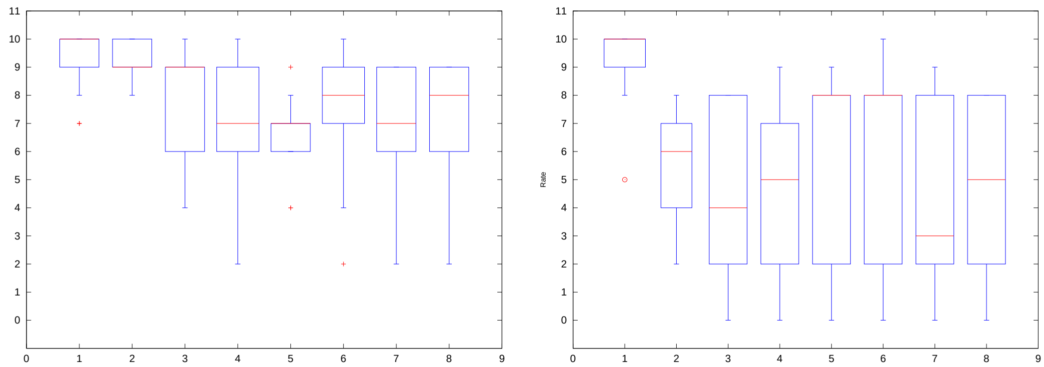

Figure 7: Box plots of the ratings of the non-expert users (left) and of the expert users (right).The numbers on the x-axis stand for the items: 1-understanding, 2-control, 3-interaction, 4-fun, 5-interest, 6-future, 7-engagement, 8-pleasure.

piece. This was also reported as a user feedback in the questionnaires. Therefore, expert users who already know how the orchestra work, may find this scenario less engaging.

4.2. Sync'n'Move evaluation

4.2.1. Participants

A group of 22 people (18 male and 3 female from different european and extra-european countries) evaluated this scenario. Mean age of the participants was 34.3 years (from 18y to 64y). The expert users group was composed by 12 people.

4.2.2. Results

Data summarized for both groups are displayed in Figure 8, whereas Figure 9 shows the results concerning expert users (left) and non-expert users (right), separately. Again, the y-axis shows the ratings range expressed as numerical values (0 for *not at all*, 10 for *very*), the x-axis shows all the items from the questionnaire *understanding*, *control*, *interaction*, *fun*, *interest*, *future*, *engagement*, *pleasure*.

As for the Mobile Orchestra Explorer scenario, also Sync'n'Move was very easy to understand

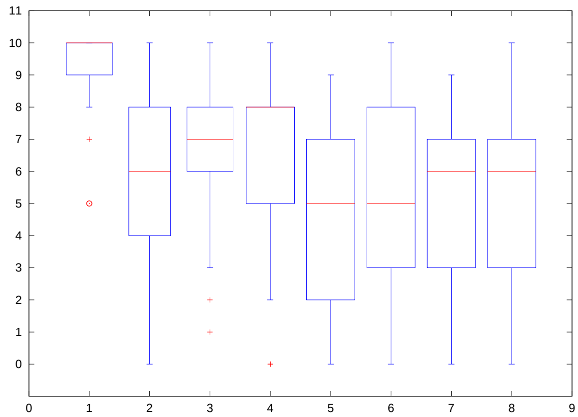

Figure 8: The global box plot

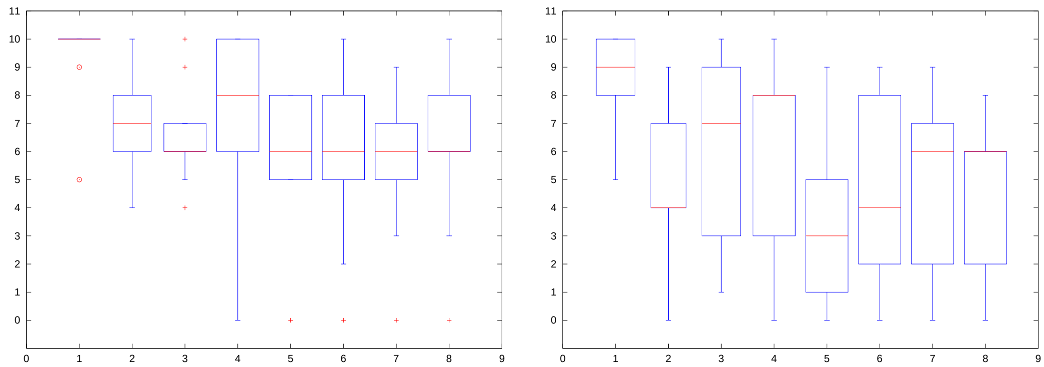

Figure 9: Box plots of the ratings of the non-expert users (left) and of the expert users (right).The numbers on the x-axis stand for the items: 1-understanding, 2-control, 3-interaction, 4-fun, 5-interest, 6-future, 7-engagement, 8-pleasure.

(median=10) and globally the ratings given by non-expert users are higher than those reported by the experts. More specifically, non-expert users answered to question about how the scenario works giving ratings higher (full 10) than those given by expert users. Non-expert users also evaluated more than satisfactory the level of *control* and *fun* during their experience. They gave neutral and low spread ratings for *interaction*, *interest*, *future*, *engagement*, and *pleasure*. Unlike, the ratings provided by expert users were more spread, for example, the middle 50% of the ranked ratings for *interaction* ranges from 3 to 9, whereas the middle 50% of the ranked ratings for *engagement* ranges from 2 to 7. We ascribe this data dispersion mainly to a greater expert users perception that the audio tempo generated by Sync'n'Move was not matching in time with the tempo chosen by the users. This could also explain the low ratings obtained for *interest* and for *future*. Moreover, Sync'n'move exploits very simple music content and interaction mechanisms, whereas expert users probably expect more complex paradigms and content.

5. CONCLUSION

This paper presents two scenarios (The Mobile Orchestra Explorer and Sync'n'Move) toward mobile applications for active experience of sound and music content exploiting KANSEI Information Processing. The scenarios were evaluated by expert and non-expert users. Feedback from users was overall positive and will be useful for improving interaction paradigms enabling a better KANSEI experience. For example, we could extend the Mobile Orchestra Explorer by allowing multiple user to explore the orchestra sections at the same time, in a collaborative way, taking into account the users behavior expressivity: if users present the same expressivity the overall orchestra playing style is modulated. For Sync'n'Move we could enable users to choose which piece of music they prefer to interact with and we could synchronize the music playback speed with the users beat.

6. ACKNOWLEDGMENTS

This work is partially supported by the FP7 EU ICT SAME Project n. 215749 on active music listening (www.sameproject.eu). The authors would like to thank Albero Massari from InfoMus Lab for creating Python scripts for reading and transmitting accelerometer data and audio streams for Nokia phones. The MAX/MSP audio spatialization plugin and the 2D spazializer speaker have been developed by IRCAM, Paris.

REFERENCES

[1] Picard, R. *Affective Computing*. MIT Press, 1997.

[2] Hashimoto, S. Kansei as the third target of information processing and related topics in japan. In *Proceedings of the International Workshop on KANSEI: The technology of emotion*, pages 101–104, 1997.

[3] `http://www.cycling74.com`.

[4] Camurri, A, Coletta, P, Varni, G, and Ghisio, S. Developing multimodal interactive systems with EyesWeb XMI. In *Proceedings of the 7th international conference on New interfaces for musical expression*, pages 305–308. ACM New York, NY, USA, 2007.

[5] Varni, G, Mancini, M, Volpe, G, and Camurri, A. Sync'n'move: social interaction based on music and gesture. In *Proceedings of the 1st international ICST conference on User Centric Media (UCMedia 2009)*. Springer LNICST, 2009.

[6] Eckmann, J. P, Kamphorst, S. O, and Ruelle, D. Recurrence plots of dynamical system. *Europhysics Letters*, 5:973–977, 1987.

[7] Zbilut, J and Jr, C. L. W. Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 5:199–203, 1992.

[8] N.Marwan, Romano, M. C, Thiel, M, and Kurths, J. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438:237–329, 2007.

[9] Romano, M, Thiel, M, Kurths, J, Kiss, I, and Hudson, J. Detection of synchronisation for non-phase coherent and non-stationarity data. *Europhysics Letters*, 71(3):466–472, 2005.