Synthesis from time-frequency statistics for environmental sound textures and music

Résumé du projet de recherche (Langue 1)

Proposition de recherche doctorale

Most of the sound texture synthesis algorithms that have been developed in recent years are based on a working definition of sound textures that has been proposed by Saint Arnaud (Saint Arnaud, 1995, Saint Arnaud et al, 1998). This definition states that sound textures are created by super position of atomic events with appearance characteristics that do not change when comparing sufficiently long segments (Saint Arnaud et al, 1998, Schwarz, 2011). Only recently due to the seminal work on texture recognition by McDermott (McDermott et al 2009, 2011) a new more perceptually oriented approach has emerged. McDermott has shown that the statistical description of the amplitude envelope in the critical bands and modulation bands of the auditory system are signal characteristics that are sufficient for the recognition of sound textures. The statistics contributing to the recognition are moments of various orders, and correlations within and between sub-bands. Based on these findings new algorithms for texture synthesis have been created (McDermott et al 2011; Liao and Roebel, 2013; Liao, 2015) that allow the synthesis by means of imposing textures statistics on noise signals. These algorithms achieve good quality when used for re-synthesis of sound textures as long as there are no sound events in the texture that are perceived individually, which means that the signal model underlying the texture analysis/re-synthesis algorithm is respected. For the moment the statistics based texture synthesis algorithms do not allow controlling the texture characteristics (e.g. modify the characteristics of a given rain sound). In this context we propose to work on the following problems based on the algorithm for texture synthesis from time frequency statistics that has initially been developed in (Liao 2015):

1. **Synthesis from statistics for environmental sound textures**

   - The possibility to synthesise from time-frequency domain statistics establishes a new synthesis paradigm that opens the possibility to create new sound spaces that are potentially highly interesting for contemporary music especially those directions that make strong use of noise sounds. Accordingly, at IRCAM, the research on sound synthesis from statistics has generated strong interest in the creative community. First experiments for musical applications have been engaged with the contemporary composer and performer Florian Hecker, who will visit IRCAM in 2016 to experiment with the new possibilities for sound synthesis. One of the objectives of the thesis is to continue collaboration with interested composers and to develop new means to create hybrid sounds for musical applications. Another potentially interesting musical application of the texture synthesis algorithm consists of synthesising dense musical sources as for examples choirs and string sections. Here again a control of texture statistics that allows manipulating the musically interesting texture properties (pitch, intensity) will need to be established.

2. **Investigation of the minimum set of correlations for the representation of a given texture signal by means of statistical parameters:**

   - The experimental results show that providing the full set of correlations the synthesised signal is over determined. As a result the result is generally reproducing the input signal. Manually limiting the maximum time shift to be taken into account for the correlation functions provides more degrees of freedom and output signals cover the space of corresponding texture signals. The present work aims to automatically find the appropriate time shifts to be taken into account when imposing auto and cross correlation functions.

3. **Investigation into controlling the properties of the synthesised sounds by means of modification of the statistics:**

   - For being able to control the texture properties by means of modification of the statistics the relation between perceived texture properties and the statistics should be established. For this the sub space within the space of the statistical description that is covered by a certain class of textures, for example rain or wind textures, should be established and within this sub space control parameters should be established that allow modifying the synthesised textures according to high level parameters (strong rain <-> weak rain). Here dimensionality reduction techniques are certainly beneficial. It should then be studied how the long-term fluctuations, in individual environmental textures (rain, wind) can be modelled. In the present implementation the long-term fluctuations are part of the texture description and can therefore not be controlled. A representation using non-stationary texture statistics would be beneficial allowing for explicit control of the time evolution of the perceived texture properties which would allow synchronising the texture evolution with its physical environment.

4. **Investigation of the minimum set of correlations for the representation of a given texture signal by means of statistical parameters:**

   - The experimental results show that providing the full set of correlations the synthesised signal is over determined. As a result the result is generally reproducing the input signal. Manually limiting the maximum time shift to be taken into account for the correlation functions provides more degrees of freedom and output signals cover the space of corresponding texture signals. The present work aims to automatically find the appropriate time shifts to be taken into account when imposing auto and cross correlation functions.

**Description of Work**

In this work we propose to control the texture characteristics (e.g. modify the characteristics of a given rain sound) by means of texture statistics that allow controlling the texture characteristics (e.g. modify the characteristics of a given rain sound).

**Introduction**

Most of the sound texture synthesis algorithms that have been developed in recent years are based on a working definition of sound textures that has been proposed by Saint Arnaud (Saint Arnaud, 1995, Saint Arnaud et al, 1998). This definition states that sound textures are created by super position of atomic events with appearance characteristics that do not change when comparing sufficiently long segments (Saint Arnaud et al, 1998, Schwarz, 2011). Only recently due to the seminal work on texture recognition by McDermott (McDermott et al 2009, 2011) a new more perceptually oriented approach has emerged. McDermott has shown that the statistical description of the amplitude envelope in the critical bands and modulation bands of the auditory system are signal characteristics that are sufficient for the recognition of sound textures. The statistics contributing to the recognition are moments of various orders, and correlations within and between sub-bands. Based on these findings new algorithms for texture synthesis have been created (McDermott et al 2011; Liao and Roebel, 2013; Liao, 2015) that allow the synthesis by means of imposing textures statistics on noise signals. These algorithms achieve good quality when used for re-synthesis of sound textures as long as there are no sound events in the texture that are perceived individually, which means that the signal model underlying the texture analysis/re-synthesis algorithm is respected. For the moment the statistics based texture synthesis algorithms do not allow controlling the texture characteristics (e.g. modify the characteristics of a given rain sound). In this context we propose to work on the following problems based on the algorithm for texture synthesis from time frequency statistics that has initially been developed in (Liao 2015):

1. **Synthesis from statistics for environmental sound textures**

   - The possibility to synthesise from time-frequency domain statistics establishes a new synthesis paradigm that opens the possibility to create new sound spaces that are potentially highly interesting for contemporary music especially those directions that make strong use of noise sounds. Accordingly, at IRCAM, the research on sound synthesis from statistics has generated strong interest in the creative community. First experiments for musical applications have been engaged with the contemporary composer and performer Florian Hecker, who will visit IRCAM in 2016 to experiment with the new possibilities for sound synthesis. One of the objectives of the thesis is to continue collaboration with interested composers and to develop new means to create hybrid sounds for musical applications. Another potentially interesting musical application of the texture synthesis algorithm consists of synthesising dense musical sources as for examples choirs and string sections. Here again a control of texture statistics that allows manipulating the musically interesting texture properties (pitch, intensity) will need to be established.

2. **Investigation of the minimum set of correlations for the representation of a given texture signal by means of statistical parameters:**

   - The experimental results show that providing the full set of correlations the synthesised signal is over determined. As a result the result is generally reproducing the input signal. Manually limiting the maximum time shift to be taken into account for the correlation functions provides more degrees of freedom and output signals cover the space of corresponding texture signals. The present work aims to automatically find the appropriate time shifts to be taken into account when imposing auto and cross correlation functions.

3. **Investigation into controlling the properties of the synthesised sounds by means of modification of the statistics:**

   - For being able to control the texture properties by means of modification of the statistics the relation between perceived texture properties and the statistics should be established. For this the sub space within the space of the statistical description that is covered by a certain class of textures, for example rain or wind textures, should be established and within this sub space control parameters should be established that allow modifying the synthesised textures according to high level parameters (strong rain <-> weak rain). Here dimensionality reduction techniques are certainly beneficial. It should then be studied how the long-term fluctuations, in individual environmental textures (rain, wind) can be modelled. In the present implementation the long-term fluctuations are part of the texture description and can therefore not be controlled. A representation using non-stationary texture statistics would be beneficial allowing for explicit control of the time evolution of the perceived texture properties which would allow synchronising the texture evolution with its physical environment.

4. **Investigation of the minimum set of correlations for the representation of a given texture signal by means of statistical parameters:**

   - The experimental results show that providing the full set of correlations the synthesised signal is over determined. As a result the result is generally reproducing the input signal. Manually limiting the maximum time shift to be taken into account for the correlation functions provides more degrees of freedom and output signals cover the space of corresponding texture signals. The present work aims to automatically find the appropriate time shifts to be taken into account when imposing auto and cross correlation functions.

**Description of Work**

In this work we propose to control the texture characteristics (e.g. modify the characteristics of a given rain sound) by means of texture statistics that allow controlling the texture characteristics (e.g. modify the characteristics of a given rain sound).
The algorithms that are developed in the thesis will open new applications in sound production for video games, film sound and music. For experimental evaluation feedback from professional users from for example the video game industry (point 2) or composers and IRCAM’s user base (point 4) should be gathered. For the experimental evaluation the algorithms need to be optimised and put into a form that allows experimentation by standard users. For ease of deployment notably to interested users within IRCAM the algorithms should be developed in python.

Informations complémentaires (Langue 1)

IRCAM is an institute bringing together researchers from many European and extra European nations. International applicants are very welcome.

Informations complémentaires (Langue 2)