Answer to the Ultimate Question of Life, the Universe, and Everything

John Smith

john.smith@umons.ac.be

Tuesday 42nd January, 2042

A dissertation submitted to the Faculty of Engineering
of the University of Mons, for the degree of Doctor of Philosophy in Engineering Science

Supervisor: Prof. J. Smith

This thesis was supported by the Funds for Meaning of Life (FML)

Jury members

Prof. **John Smith** - Université de Mons, President
Prof. **John Smith** - Université de Mons, Supervisor
Prof. **John Smith** - Université de Mons, Co-supervisor
Dr. **John Smith** - Université de Mons
Dr. **John Smith** - John Smith & co

"Forty-two,"
said Deep Thought, with infinite majesty and calm.

Acknowledgements

I want to thank the people.

Last but not least, I cannot skip this page without mentioning my supervisor.

Contents

List of acronyms

Chapter 1

Laughter databases

Contents

This chapter is based on the following publication :

- "*The AV-LASYN database : A synchronous corpus of audio and 3D facial marker data for audiovisual laughter synthesis*", in Proc. of the 9th Int. Conf. on Language Resources and Evaluation (LREC), 2014.

To be able to analyse, model and synthesize laughter, samples are needed. In this chapter, we will pesent the Audio-Visual LAughter SYNthesis (AV-LASYN) database in details, as it is a database recorded in the framework of this PhD thesis and was used for the developments made in this scope. Other databases containing laughter exist but do not include facial motion capture data except for the AVLaughterCycle (AVLC) database which will be briefly described below and therefore are out of the scope of this work. A comprehensive list of databases in which laughter has been spotted may be found in [1].

This chapter is organised as follows : Section **??** presents the AVLC database, Section 1.1 presents the AVLASYN database, Section **??** details a Principal Component Analysis (PCA) on the AVLASYN database, Section **??** explains how the data was segmented and annotated and Section **??** describes the content of the database and the chapter is concluded by a brief summary and possible perspectives.

1.1 The AV-LASYN database

A new database was built in the scope of this thesis. Two main reasons motivated the recording of the AV-LASYN database :

1. The only available laughter database that includes facial motion capture data was the AVLC database and, as stated in the previous section, it needed a significant amount of post-processing to be ready for research.
2. Even if the data from AVLC database was cleaned and ready, the amount of data for one given subject was too small for the aim of this thesis which is building models for synthesis (among the 24 subjects in AVLC database, the subject with the highest number of expressive laughs has around 3 minutes of data).

It was thus more relevant to build a new database with as much data as possible for one given subject and with clean facial motion capture data. A synchronous database of acoustic and 3D facial marker data was built for audiovisual laughter synthesis. Since the aim is to use this database for Hidden Markov Model (HMM)-based modelling and synthesis, the amount of collected data from one given subject had to be maximized. The corpus contains 250

utterances of laughter from one male participant. Laughter was elicited with the help of humorous videos. The resulting database is synchronous between modalities (audio and 3D facial motion capture data). Visual 3D data is available in common formats such as BVH and C3D [1] with head motion and facial deformation independently available. Data is segmented and audio has been annotated. Phonetic transcriptions are available in the HMM Toolkit (HTK)-compatible format. A PCA [2] has been conducted on visual data and has shown that a dimensionality reduction might be relevant. The following sections will give details on these different aspects.

1.1.1 Recording Pipeline

This section gives information about the experimental setup used for recordings. Figure 1.1 gives an overview of the recording pipeline.

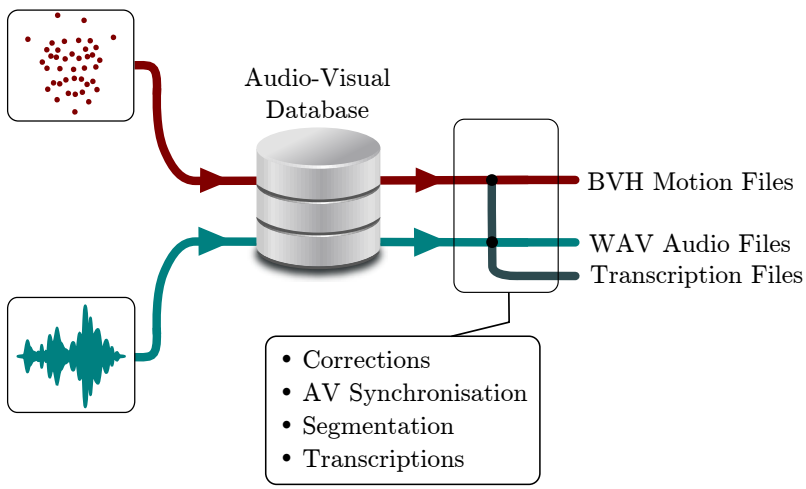

Figure 1.1. Data recording pipeline

[1] These formats are explained later in the chapter.

The Stimuli

The laughs were triggered by humorous videos found on the Internet. The subject was free to watch whatever he could find as far as it was funny for him. A total amount of 125 minutes were watched by the subject to build this corpus.

1.2 In brief

Summary of Chapter 1

- In this chapter, we have shown a recording protocol and the fundamental post-processing steps to follow in order to prepare data for audiovisual laughter synthesis.
- The PCA on visual data confirmed our thoughts that 99 dimensions (33 markers times 3 coordinates) may be reduced to describe the facial deformation during laughter with a lower number of dimensions.
- This does not mean that we do not need 33 markers but that the motions of these markers are related to each other and therefore a dimensionality reduction may be applied to data while still correctly representing the motion as shown in Chapter **??**.

Perspective for Chapter 1

- Although the collected database is the biggest known for a single subject with both audio and facial motion capture data, it is still quite small. The most difficult part in the database building process is the manual annotation part. A very valuable future work would be the development of an automatic laughter phone recognizer to build phonemic transcription from the audio files. This would allow to augment the size of the database drastically and relatively easily.

Appendix A

Publications related to this thesis

A.1 Regular Papers (Referenced by Scopus)

1. H. **CAKMAK**, K. EL HADDAD and T. DUTOIT, "*Collecting, analyzing and estimating laughter arousal*", IEEE International Conference on Multimedia and Expo (ICME), Seattle, USA, 2016, [under review]
2. H. **CAKMAK**, K. EL HADDAD, P. MARIGHETTO, B. TURKER, H. KHAKI, S. MARZBAN, J. LEROY, N. RICHE, "*A synchronous multimodal emotion database*", IEEE International Conference on Multimedia and Expo (ICME), Seattle, USA, 2016, [under review]
3. H. **CAKMAK**, K. EL HADDAD and T. DUTOIT, "*Laughter Arousal Estimation From Acoustic Features*", [ready for submission]

A.2 Papers in International Conference with Peer Review

A.3 Book chapters

Bibliography

[1] J. Urbain, "Acoustic laughter processing", Ph.D. dissertation, University of Mons, 2014.

[2] I. T. Jollife, *Principal Component Analysis*, October 2002.

List of Figures

List of Tables

This thesis was made using a customized version of “hepthesis”.