

2008-01-01

Emotional Speech Corpora for Analysis and Media Production

Charlie Cullen

Technological University Dublin, charlie.cullen@tudublin.ie

Brian Vaughan

Technological University Dublin, brian.vaughan@tudublin.ie

Spyros Kousidis

Technological University Dublin, spyros.kousidis@tudublin.ie

Follow this and additional works at: <https://arrow.tudublin.ie/dmcccon>



Part of the [Other Computer Engineering Commons](#)

Recommended Citation

Cullen, C., Vaughan, B. & Kousidis, S. (2008) Emotional Speech Corpora for Analysis and Media Production. *SAMT, 3rd International Conference on Semantic and Digital Media Technologies*, Koblenz, Germany, 3-5 December.

This Conference Paper is brought to you for free and open access by the Digital Media Centre at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.

Emotional Speech Corpora for Analysis and Media Production

Charlie Cullen, Brian Vaughan, Spyros Kousidis¹
and John McAuley¹

¹ Digital Media Centre, Dublin Institute of Technology, Aungier Street,
Dublin 2, Ireland
{charlie.cullen, brian.vuaghan, spyros.kousidis, john.mcauley}@dmc.dit.ie

Abstract. Research into the acoustic correlates of emotional speech as part of the SALERO project has led to the construction of high quality emotional speech corpora, which contain both IMDI metadata and acoustic analysis data for each asset. Research into semi-automated, re-usable character animation has considered the development of online workflows based on speech corpus assets that would provide a single point of origin for character animation in media production. In this paper, a brief description of the corpus design and construction is given. Further, a prototype workflow for semi-automated emotional character animation is also provided, alongside a description of current and future work.

Keywords: Emotional Speech Corpora, Emotional Character Animation, Speech Analysis.

1 Introduction

Speech assets form a crucial part of many media productions, yet the proper analysis and metadata tagging of such assets is often overlooked as a similarly crucial aspect of the production workflow. Current research undertaken as part of the EU Salero project [1] seeks to produce intelligent reusable media content, and to this end research in emotional speech analysis has been considered relative to media production workflows. This work has highlighted both the need for metadata tagging in media production and also the current lack of effective implementation of tagging based workflows. For example, in a typical character animation workflow a speech asset is analysed manually (by scrubbing the timeline) or semi-automatically (by correcting output from automated lip-synching plugins) for each animation instance in turn, with none of the analysis performed being retained for subsequent re-use. In such situations, a change of character or context would dictate a completely new workflow, with none of the previous information being retained.

2 Emotional Speech Corpora Design and Construction

The emotional speech corpus used [2, 3] is built on MySQL, with automated annotation tools being implemented in Ajax and Ruby on Rails to allow batch upload and metadata annotation of assets to be performed. Acoustic and emotional analysis data relating to vowels in each asset is obtained by batch analysis [4], and is held in

standard SMIL format for querying. Each asset is also annotated using the IMDI metadata [5] schema, which defines a speech act in relative groups as described in Table 1:

Table 1. IMDI schema breakdown. Each grouping within the schema is briefly described.

IMDI Grouping	Description
Project	Logical organization of recorded sessions
Session	A particular recording containing speech acts
Actor	A participant in a recording session
Content	Definition of genre, sub-genre, e.t.c.

3 Prototype Character Animation Workflow

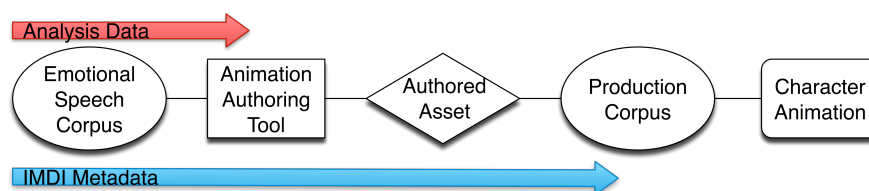


Fig. 1. The diagram above demonstrates the use of corpus data throughout the animation workflow. The analysis data relating to emotional and acoustic parameters is used by the authoring tool, while IMDI metadata is retained by an asset throughout. An authored asset is stored in a separate corpus, for use in re-usable character animation processes.

References

1. Haas, W., Thallinger, G., Cano, P., Cullen, C., Bärger, T.: SALERO - Semantic Audiovisual Entertainment Reusable Objects. International Conference on Semantic and Digital Media Technologies (SAMT), Athens, Greece (2006)
2. Cullen, C., Vaughan, B., Kousidis, S., Wang, Yi., McDonnell, C. and Campbell, D. : Generation of High Quality Audio Natural Emotional Speech Corpus using Task Based Mood Induction International Conference on Multidisciplinary Information Sciences and Technologies Extremadura, Merida (2006)
3. Vaughan, B., Kousidis, S., Cullen, C.: Task-Based Mood Induction Procedures for the Elicitation of Natural Emotional Responses. The 5th International Conference on Computing, Communications and Control Technologies: CCCT 2007, Orlando, Florida, USA (2007)
4. Cullen, C., Vaughan, B., Kousidis, S., Reilly, F.: A vowel-stress emotional speech analysis method. The 5th International Conference on Cybernetics and Information Technologies, Systems and Applications, CITSA 2008, Genoa, Italy (2008)
5. ISLE: IMDI (ISLE Metadata Initiative), Metadata Elements for Session Descriptions. (2003)