2009-01-01

CorpVis: an Online Emotional Speech Corpora Visualisation Interface

Charlie Cullen
Technological University Dublin, charlie.cullen@tudublin.ie

Brian Vaughan
Technological University Dublin, brian.vaughan@tudublin.ie

CorpVis: An Online Emotional Speech Corpora Visualisation Interface

Charlie Cullen[1], Brian Vaughan[1], John McAuley[1], Evin McArthy[1]

[1] Digital Media Centre, Dublin Institute of Technology, Aungier Street, Dublin, Ireland
charlie.cullen@dmc.dit.ie

Abstract. Our research in emotional speech analysis has led to the construction of several dedicated high quality, online corpora of natural emotional speech assets. The requirements for querying, retrieval and organization of assets based on both their metadata descriptors and their analysis data led to the construction of a suitable interface for data visualization and corpus management. The CorpVis interface is intended to assist collaborative work between several speech research groups working with us in this area, allowing online collaboration and distribution of assets to be performed. This paper details the current CorpVis interface into our corpora, and the work performed to achieve this.

Keywords: Emotional speech, online speech corpora, data visualization

1 Introduction

Existing emotional speech corpora are often maintained offline, with interface and visualization tools often being developed on an individual basis if at all. Although useful work has been performed on multidimensional scaling analysis [1, 2] and tool development [3-5], no online solution for emotional speech corpora visualization currently exists. Online data visualization is a growing field of research with examples such as Google Finance [6], Amazon [7] and Flickr [8] becoming a popular means of providing interactive user interfaces into large, real-time data sources. This suggests a useful means of implementing a similarly scalable and reusable data visualization toolset for emotional speech corpora. Our team has used these components to form the basis of an online emotional speech corpora interface- CorpVis. The CorpVis interface is designed to allow speech researchers to visualize corpora assets, query the visualization in real time and view the results in detail at asset level.

2 Emotional Speech Corpora

Work in the SALERO project [9] has developed several corpora of high quality emotional speech assets [10, 11]. These corpora form the basis of work in emotional speech analysis and synthesis, linguistic convergence and machine learning. The only

defined attempt at corpus metadata standardization performed thus far the is the ISLE Metadata Initiative (IMDI) [12].

In IMDI, separate session bundles are grouped logically under an overarching project. Each session in turn relates to a specific type of content, and involves various actors who produce the speech assets that are analyzed for acoustic, linguistic and emotional information. The hierarchical groupings used by IMDI often contain complex sets of descriptors, with redundancy of these descriptors being a difficult issue to resolve when seeking to avoid large amounts of effort in the cataloguing of speech assets. For this reason, elements of the IMDI schema are suppressed by the CorpVis interface, to ensure the overall data visualization is kept as simple as possible.

Each asset in the corpus is analyzed for acoustic data including pitch, intensity, contour [13] and voice quality information such as jitter and shimmer [14]. In addition, manual annotation of linguistic information (such as IPA transcription) can also be stored with the analysis data within a standard SMIL format XML file. Emotional data [15] is also included, though rating of emotions requires group listening tests [16] to validate each result.

3 CorpVis Visualization Implementation

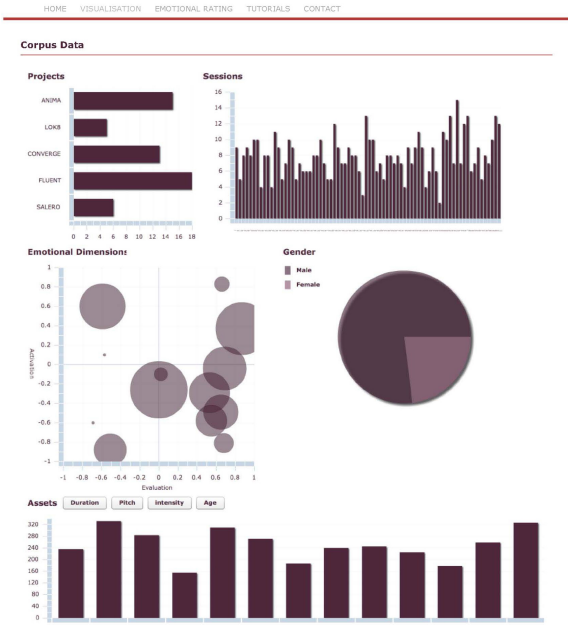

Fig. 1. Screen shot of the CorpVis corpora visualisation tool. *Tutorials on how to use the tool and what it does are provided for the user on the DMC website.*

The CorpVis application is divided into three separate tiers: a presentation tier developed using Adobe Flex; an application tier developed using Ruby on Rails, with

support for SMIL through the REXML ruby-gem and a data tier consisting of a MySQL database. The presentation tier implements the corpus visualization. The visualization tool delivers project, session, asset and emotional dimension information in a single interactive screen. The user can also visualize gender, age and acoustic data relating to each asset grouping. If analysis of an individual asset is required, a separate asset analysis screen is launched (Figure 2):

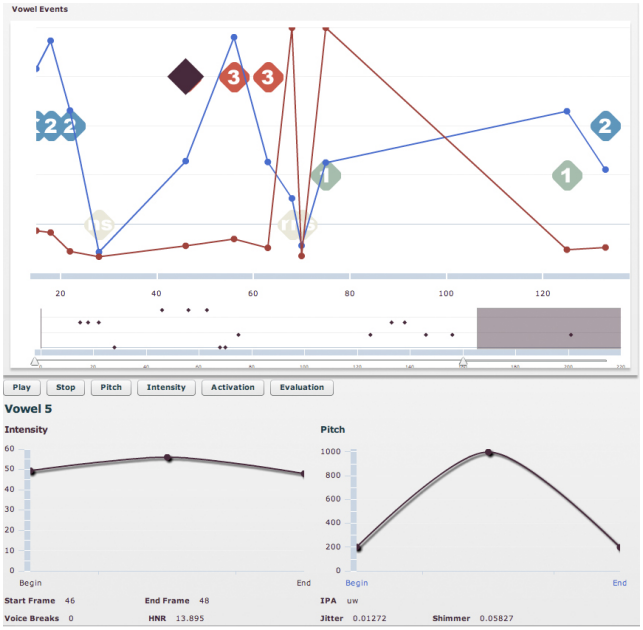

Fig. 2. Screen shot of the CorpVis asset viewer tool. *This tool provides a visualization of all acoustic, linguistic and emotional data relating to a single asset. The vowels in a speech act and their prominence are shown in the top chart alongside pitch, intensity and emotional dimension curves. The bottom graphs show pitch, intensity and voice quality for the selected vowel. The user can also add annotations (not shown) that will update the corpora database.*

A full demonstration of the CorpVis interface is available online at www.dmc.dit.ie. Due to ethical considerations in emotional speech research, full access can be obtained by directly contacting members of the research team.

4 Ongoing and Future Work

This paper details a brief introduction to the CorpVis emotional speech corpora visualisation tool. Development is ongoing, aiming to further streamline the operation of the interface and provide more flexible visualization and querying options. The implementation of automated emotional and linguistic analysis is also currently being developed, leveraging machine learning algorithms to develop a completely automatic analysis method for all speech assets in the corpus.

Acknowledgments. The research leading to this paper was partially supported by the European Commission under contract IST-FP6-027122 “SALERO”.

References

[1] K. Yamakawa, T. Matsui, and S. Itahashi, "MDS-based Visualization Method for Multiple Speech Corpus Features," *IEICE Technical Report,* vol. 108, pp. 35-40, 2008.

[2] A. Batliner, S. Steidl, C. Hacker, and E. Nöth, "Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech," *User Modeling and User-Adapted Interaction,* vol. 18, pp. 175-206, 2008.

[3] K. Sjölander and J. Beskow, "Wavesurfer - an open source speech tool," in *Sixth International Conference on Spoken Language Processing (ICSLP 2000)*. vol. 4 Beijing, China, 2000, pp. 464-467.

[4] R. Kubat, P. DeCamp, and B. Roy, "TotalRecall: Visualization and Semi-Automatic Annotation of Very Large Audio-Visual Corpora," in *Proceedings of the 9th international conference on Multimodal interfaces*, A. Nagoya, Ed., 2007, pp. 208-215.

[5] C. Barras, E. Geoffroisb, Z. Wuc, and M. Libermanc, "Transcriber: Development and use of a tool for assisting speech corpora production," *Speech Communication,* vol. 33, pp. 5-22, 2001.

[6] B. Consulting, "Google Finance Visualisation Tool," Webmaster, 2008.

[7] S. Bannur, "Querying Amazon through webservice," 2008.

[8] F. T. Research and D. LLC, "Pikeo- share your world, explore another," 2008.

[9] W. Haas, G. Thallinger, P. Cano, C. Cullen, and T. Bu¨rger, "SALERO - Semantic Audiovisual Entertainment Reusable Objects," in *International Conference on Semantic and Digital Media Technologies (SAMT)* Athens, Greece, 2006.

[10] C. Cullen, B. Vaughan, and S. Kousidis, "Emotional speech corpus construction, annotation and distribution," in *The sixth international conference on Language Resources and Evaluation, LREC 2008* Marrakech, Morocco, 2008.

[11] C. Cullen, 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 " in *International Conference on Multidisciplinary Information Sciences and Technologies* Extremadura, Merida, 2006.

[12] ISLE, "IMDI (ISLE Metadata Initiative), Metadata Elements for Session Descriptions," Draft Proposal Version 3.0.3 ed, 2003.

[13] C. Cullen, B. Vaughan, S. Kousidis, and F. Reilly, "A vowel-stress emotional speech analysis method," in *The 5th International Conference on Cybernetics and Information Technologies, Systems and Applications, CITSA 2008* Genoa, Italy, 2007.

[14] A. Ozdas, Shiavi, R.G., Silverman, S.E., Silverman, M.K., Wilkes, D.M., "Investigation of vocal jitter and glottal flow spectrum as possible cues for depression and near-term suicidal risk," *IEEE Biomedical Engineering,,* vol. 51, pp. 1530 - 1540, 2004.

[15] R. Cowie and R. R. Cornelius, "Describing the emotional states that are expressed in speech," *Speech Communication Special Issue on Speech and Emotion,* vol. 40, pp. 5-32, 2003.

[16] B. Vaughan and C. Cullen, "Emotional Speech Corpus Creation, Structure, Distribution and Re-Use," in *Young Researchers Workshop in Speech Technology (YRWST2009)* Dublin, Ireland, 2009.