GALA 2008 submission document

Title:	Donald, a speech-driven Character
Track:	Student
Category:	Tool
Authors:	Gregor Hofer
	3rd year PhD Student
	CSTR, University of Edinburgh
	Junichi Yamagishi
	Researcher
	CSTR, University of Edinburgh

Contact data: **name: Gregor Hofer**
e-mail: g.hofer@sms.ed.ac.uk
url: *http://homepages.inf.ed.ac.uk/s0343879/*
phone:
postal address: *Affiliation (if appl).*
Street/PBox, Number, Town, State(if appl), Country

URL: *http://homepages.inf.ed.ac.uk/s0343879/*

Movie file submitted: *GregorHofer.mov*

Reference teacher: Hiroshi Shimodaira
affiliation: University of Edinburgh
e-mail: h.shimodaira@ed.ac.uk

Process of development: Donald was developed at the Centre for Speech Technology Research as part of my PhD work. The research is concerned with automatically driving a talking head from speech. To demonstrate the capabilities of the developed models, Donald was designed and implemented using Blender and python scripting. The 2nd Author is one of the developers of HTS.

Resources used: 1. Blender, an open source 3D modeling and animation package, was used for control and rendering of the character.
2. HTK, a Hidden Markov Toolkit, was used to train the developed models.
3. HTS, an extension of HTK, was used to generate trajectories from the models

Resources required: Donald can be run on any machine that can run Blender. It has been successfully tested on Windows, Mac Os X, Linux. A OpenGL compatible Graphics Card is recommended for decent real time performance.

GALA 2008 submission document

Title Donald, a speech-driven Character

1. The application and context of the work
The developed tool, is intended to be used in conjunction with a talking head. Although it has not been used in a wider context yet, it could easily be integrated into any application. As the input is just speech, the system can be deployed whenever the input text is unknown, and just the speech signal is available.

2. Novelty
Our system is completely speech-driven and does not require text input. In addition to lip motion, the feasibility of generating other kinds of motion has been demonstrated by means of head motion. The models are trained on motion capture data and can be adapted to any speaker. Most current systems produce deterministic output, meaning that the talking head is always behaving the same, when saying the same sentence. To emulate the phenomena that humans move and express themselves in many different ways while saying the same sentence over and over again, stochastic output can also be generated.

3. The architecture
The tool that controls Donald is implemented as a command line application that outputs files, which can be used to drive blend shapes and bones of a 3D model. The tool has options for controlling the dynamic range of the generated trajectories and an option to turn on stochastic or deterministic output.

4. Performance
The command line tool produces output in near real time. The submitted videos are rendered offline but it is possible to drive a realtime talking head if wanted.

References (optional)

BUSSO, C., DENG, Z., NEUMANN, U., AND NARAYANAN., S. 2007. Rigid head motion in expressive speech animation: Analysis and synthesis. IEEE Trans. Audio, Speech and Language Processing 15, 3, 1075–1086.

BRAND, M. 1999. Voice puppetry. In Proceedings of the 26th annual conference on Computer graphics and interactive techniques (SIGGRAPH'99), 21–28.

TOKUDA, K., YOSHIMURA, T., MASUKO, T., KOBAYASHI, T., AND KITAMURA, T. 2000. Speech parameter generation algorigthms for HMM-based speech synthesis. In Proc. ICASSP 2000, 1315–1318.