Designs that talk and listen: Integrating functional information using voice-enabled CAD systems

Amar Kumar Behera, Alison McKay
School of Mechanical Engineering
University of Leeds
Leeds, UK
a.k.behera@leeds.ac.uk, a.mckay@leeds.ac.uk

***Abstract*— In this paper, we report a novel approach to voice-enabled design synthesis and management. This approach consists of integrating "text-to-speech" and "speech-to-text" tools within a software prototype developed in MATLAB that connects seamlessly to a computer aided design software via an application program interface. Users can train the software to recognize specific words spoken by them and subsequently query for functions of individual parts or sub-assemblies. The results indicate the feasibility of interactive voice enabled product lifecycle management (PLM) tools in the future, where designing on the fly can be increasingly supported.**

Keywords—voice-enabled CAD; text-to-speech; speech-to-text; function; PLM

I. Introduction

The next generation of collaborative and design-in-motion solutions calls for the integration of advanced vision and speech processing into current computer aided design (CAD) packages. Efforts have been made to enable the same, which include interfaces such as Speak4CAD [1] and Natural Voice-Enabled CAD [1, 2], and gesture recognition [3] and eye-tracking tools [4]. While voice-based systems are on the rise, their popularity in the engineering design has been limited. Part of this can be attributed to the lack of integration of part and product functionality into such systems. In this work, an effort is made to develop a prototype where a user can interactively query for part functionality in CAD models.

II. Related Work

A. *Voice-enabled CAD systems*

Current computer aided design systems typically involve a large number of functionalities available to users as menu items within the graphical user interface (GUI) that enable the creation and manipulation of design definitions such as 3D models. However, their use usually requires training users to move around the interface and use these menu items intelligently using mouse and keyboard interactions to realize the desired design definitions. Consequently, their ease of use is limited thereby reducing productivity in product development firms. To alleviate these limitations, a number of other techniques of human-computer interaction have been proposed. These include haptic interfaces [5], freehand drawing [6], gesture recognition [3], eye tracking [4] and voice integration [1]. Of these, voice integration is the focus of this work.

A summary of voice-enabled CAD systems and their key contributions is presented in Table 1. It may be noted that almost all of this work deals with the creation and manipulation of 3D CAD models. The objective of the present work, on the other hand, is to illustrate the integration of functional information into CAD models using voice. This enables designers and managers to interactively query complex 3D CAD models for functional information. This, in turn, has the potential to assist them with understanding the function of individual parts and sub-assemblies as well as the overall function of the product, and to support changes based on this information.

TABLE I. Voice Integration With CAD

Authors	**Contribution**	**Reference**
Gao	Voice commands used together with 3D input device to create, manipulate and visualize 3D objects	[7]
Chu	Multi-sensory interface consisting of voice input, hand motion tracking and auditory output	[4]
Weyrich	Virtual workbench with 6D input device and holographic projection of 3D scene	[8]
Bolt	Converting speech and tracking of both hands into input concepts for model generation	[9]
Kou	Natural language conversations processed for voice-enabled CAD	[2, 10]
Dani	Bi-modal hand and voice tracking	[11]

B. *Function: modeling, representation and modification*

Function is used in engineering design to represent the relationship between the inputs and outputs when a product is in operation [12]. Often, an overall complex function can be broken down into sub-functions of lower complexity. These sub-functions can then be used to find design solutions and also be combined together to form function structures.

Approaches to computational design synthesis using function, structure and behavior combined with graph grammars are now available [13, 14]. A standardized representation of function has also been proposed that involves the use of a function information model that consists of separate function schema and flow schema [15]. Other frameworks that been proposed include a DSM-based approach [16], using divergent-convergent analysis of concepts [17], using a functional basis based on verb-object format [18], using qualitative data analysis tools [19] etc. The use of different function models in interdisciplinary design across 10 different engineering firms has also been recently surveyed [20]. Although different approaches and models are available for representing and using function, lucid methods of representing function in a natural way that are easily applied in industry are still lacking. To alleviate this need, in this paper, we propose a new method that uses voice to integrate functional information in design definitions that can then be retrieved by a user by simply interacting by talking with the design.

III. METHODOLOGY

A software package (IVEPLM) was built in MATLAB that consists of three separate elements, as illustrated in Fig. **1**. These are: i) a speech recognition module, ii) a text-to-speech converter and iii) a live connection between MATLAB and a CAD software package, SolidWorks. Speech recognition and text-to-speech conversion were realized by carrying out feature extraction, acoustic matching and subsequent decoding with the help of custom algorithms built around a MATLAB toolbox, VOICEBOX [21]. The live connection with SolidWorks was realized with the help of a toolbox, CADLab [22]. Details of how the 3 separate elements of the developed software work are discussed in the following sub-sections.

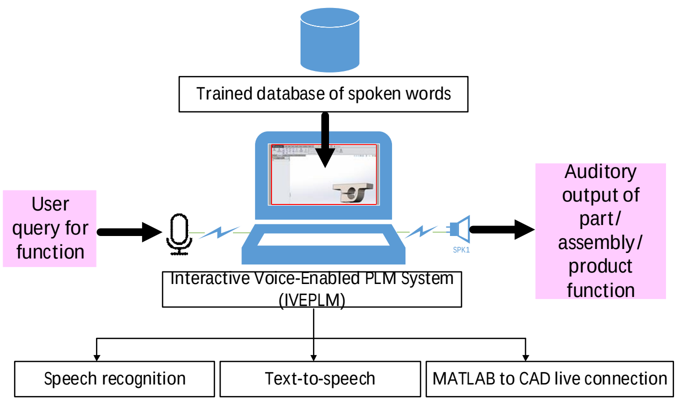


Fig. 1. Schematic of the voice-enabled function query system for next-generation product lifecycle management (PLM)

A. Speech recognition

Speech recognition can typically be of three types of speech: i) isolated words, ii) continuous words with a restricted vocabulary and iii) large vocabulary continuous speech recognition. Typical steps in the speech recognition process involve discarding components that are not essential to the recognition (e.g. pitch, mean value and phase), spectral analysis, framing, acoustic matching and decoding. Spectral analysis was carried out using mel-frequency cepstral coefficients (MFCC) obtained by linear predictive coding (LPC) analysis, as discussed by Gudnason et al. [23]. Acoustic matching was then performed by calculating the distance between two vectors, one for the voice signal of the user of IVEPLM and the other a training set of vectors. This distance was calculated while performing dynamic time warping (DTW) on the two vectors so as to account for difference of timing that may occur when the same words are spoken with different speeds. The vector with the minimum distance to the trained set was then recognized as the words of the user.

B. Text-to-speech conversion

Text-to-speech conversion is carried out using the Microsoft Win32 Speech API (SAPI), which has been implemented within a sub-routine, sapisynth.m [24]. Sapisynth is an advanced version of an original speech synthesizer written by Deng [25]. The output voice attributes from this sub-routine can be adjusted to reflect the voice of a person with a specific gender and age.

C. Live connection between MATLAB and SolidWorks

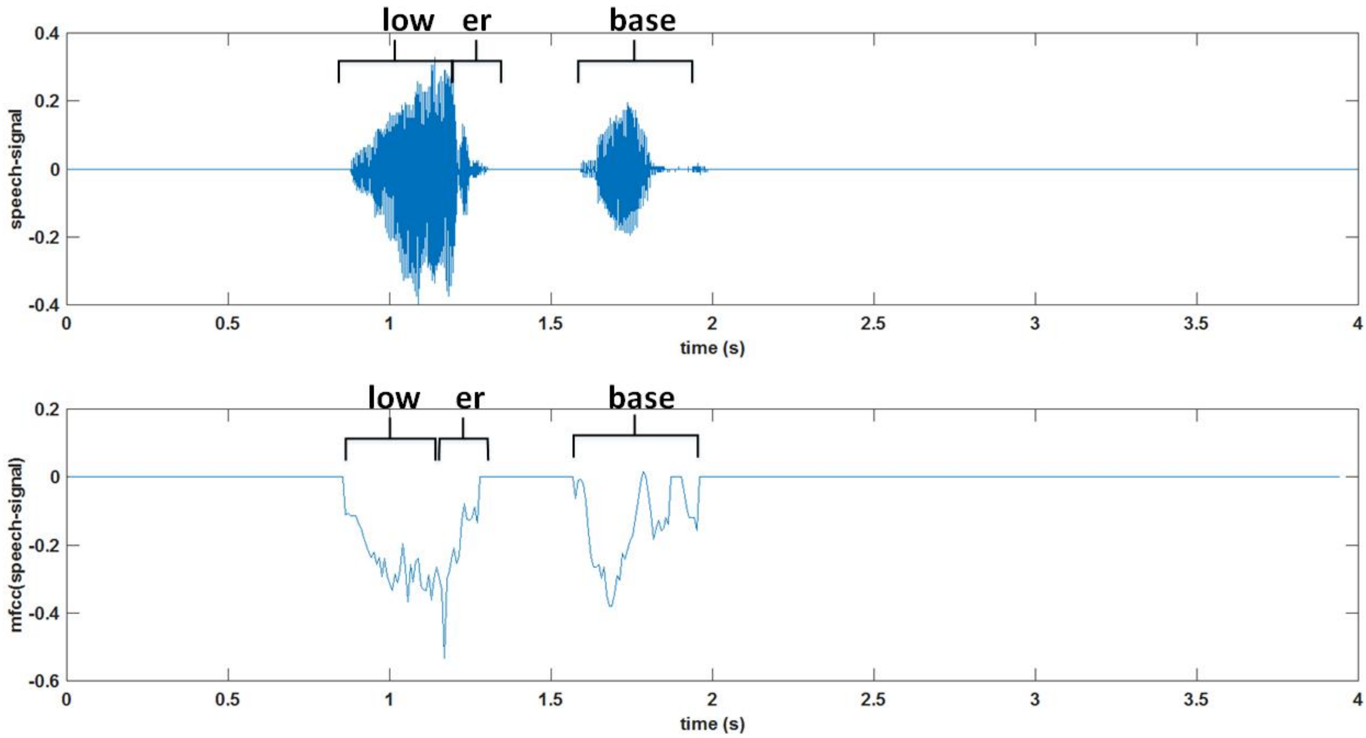
The live connection between MATLAB and SolidWorks is realized using a CAD toolbox for MATLAB [22]. This toolbox enables an user to access SolidWorks part, assembly or drawing data as variables within MATLAB. Any changes made to the SolidWorks model are reflected in MATLAB, while commands from MATLAB get implemented within the SolidWorks interface. It also enables the user to carry out solid modeling such as extruding, cutting, and revolving from within the MATLAB environment.

IV. RESULTS

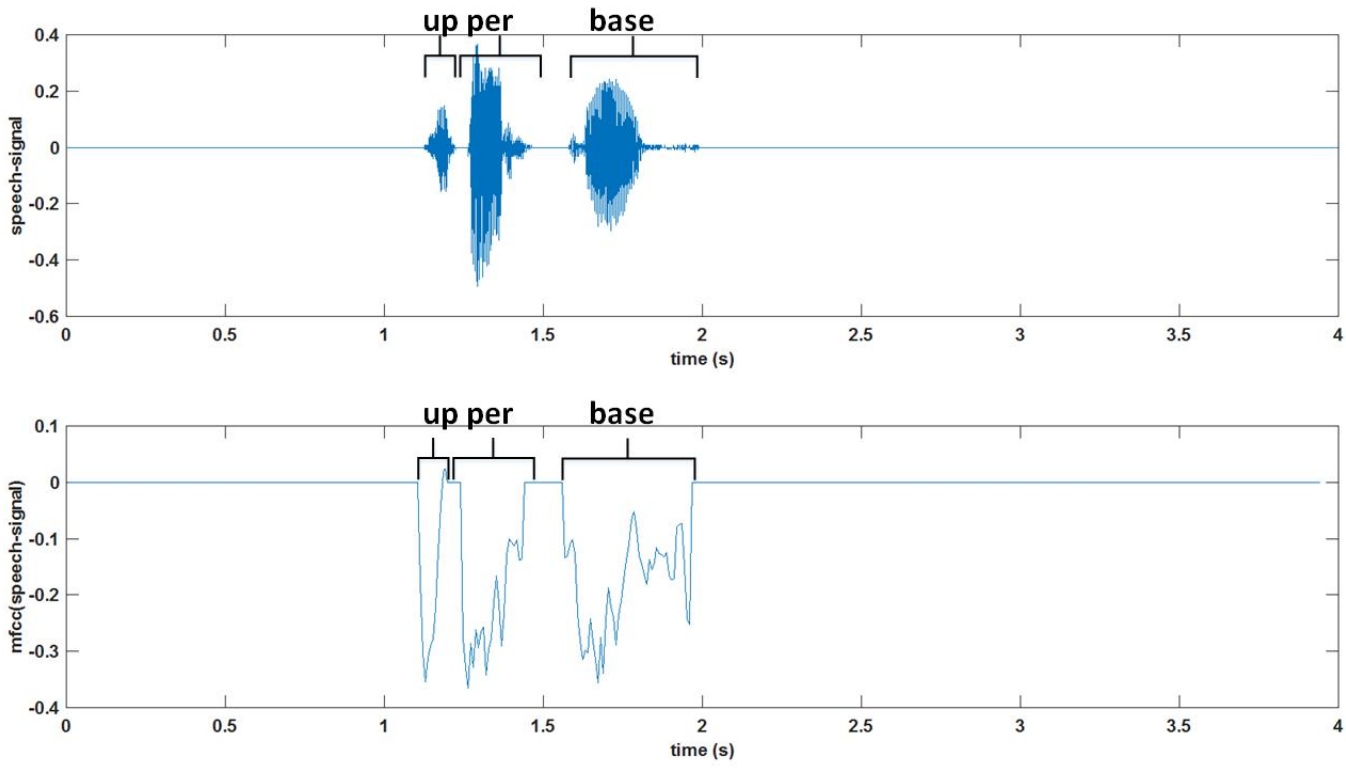
The results from the three elements that form part of the developed approach to functional integration are presented below. Tests were carried out on different CAD models; however, a specific case of a robotic arm is used to present the results.

A. Speech recognition results

The speech signal was processed to observe the effect of finding the mel-frequency cepstral coefficients for different words. Fig. 2 shows the results for two words ‘lower base’ and ‘upper base’. It was noted that the word ‘lower’ has two syllables and is pronounced as ‘low-er’, while the word ‘upper’ has two syllables and pronounced as ‘up-per’ and the word ‘base’ consists of a single syllable. Corresponding to these syllables, one can see three dips in each MFCC plot. However, as the syllable ‘up’ is shorter than the syllable ‘low’, the two dips in ‘upper’ are very distinctly separated from one another. Also, as there is a pause between ‘lower’ and ‘base’, and between ‘upper’ and ‘base’, this is also clearly reflected in the MFCC results.



(a)



(b)

Fig. 2. Speech signal and MFCC for the words (a) lower base (b) upper base

These experiments were followed by detection experiments on the words for the different parts of the robotic arm to find out the misclassification rate. 20 repeats were performed for each word. These results are presented in Table 2 alongwith maximum and minimum vector distances of detected words from the training set. The words ‘lower’, ‘upper’, ‘base’ and ‘arm’ are common across 2 words each and hence, these part names were ideal for testing the consistency of the acoustic matching algorithms performed by carrying out dynamic time warping. It may be noted that the misclassification rate for ‘grasper assembly’ and ‘robot’ was much lower than for the remaining parts due to their distinctiveness from the other words. However, ‘upper base’ was an exception to this.

TABLE II. MISCLASSIFICATION RATE FOR DIFFERENT PARTS

Part Name	Misclassification rate (%)	Minimum distance	Maximum distance
Lower base	15	0.8524	1.8987
Upper base	0	1.5104	3.8809
Lower arm	40	1.4141	4.5295
Upper arm	10	1.5276	3.7346
Grasper assembly	0	2.0891	4.6019
Robot	5	0.5930	1.8814

B. *Text-to-speech results*

The results for text-to-speech conversion were nearly perfect, as far as converting text to the appropriate word sounds was concerned. Different functions of the individual parts were processed as full sentences by the speech synthesizer. Examples of these sentences are provided in Table 3. Transitions from one sentence to another were not always very well handled. Besides, the speech synthesizer does not quite communicate naturally like humans with pause and emphasis as needed. Rather, it combines the stored signals from different words to form sentences, which sometimes give the user an artificial feel.

TABLE III. SYNTHESIS OF PART, ASSEMBLY AND PRODUCT FUNCTIONS AS SPEECH

Part	Function
Lower base	This part supports the entire robot.
Upper base	The upper base connect the lower base to the lower arm and allows for rotation of the lower arm.
Lower arm	The function of this part is to connect the upper arm and grasper assembly to the base and allow their movement.
Upper arm	The upper arm holds the grasper assembly.
Grasper assembly	This part holds the grasper and moves it to grip objects.
Robot	The robot has five main components. These are used together to manipulate different objects of interest.

C. *Live connection results*

A snapshot of a live connection within the IVEPLM interface is shown in Fig. 3. On running the interface the user is asked by the interface: “Name the part whose function you wish to learn”. In response, the user responds with a part, assembly or product name, which is then recognized by the software and the interface responds with the function as an audible sentence. Changes made to the SolidWorks model are reflected within the MATLAB interface immediately. For instance, a new hole was extruded in the robot, and consequently, the mass of the robot reflected in MATLAB got immediately updated.

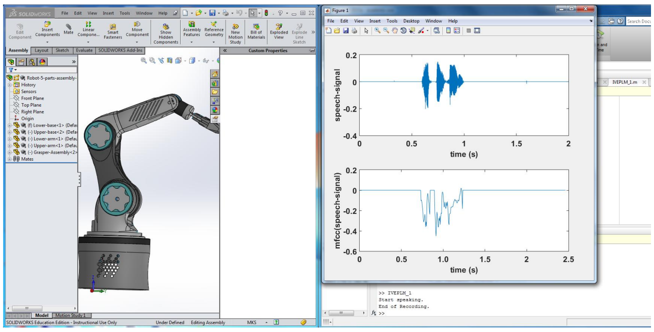


Fig. 3. A snapshot of the IVEPLM interface showing a live connection between MATLAB and SolidWorks with voice integration

V. DISCUSSION

Functional integration with CAD has been long researched as important problem in the design process as well as subsequent product lifecycle management. By integrating function in a systematic manner with voice, futuristic PLM tools can enable engineers and managers involved in the product lifecycle to make design decisions quickly, implement design changes in a robust manner and learn about the design quickly. With respect to the voice-enabled CAD summarized in Table I, the novelty of this research is two-fold: the integration of functional information into CAD and the use of voice control to query functional information from CAD. Prior work, as in Gao et al. [7], Bolt et al. [9] and Kou et al. [2, 10] has mainly focused on creating CAD models using voice without considering integration with design structures such as bills of materials, function structures and assembly mating conditions, which are critical to the use of product definitions in an industrial application context. This paper introduces the next level of integration of voice by incorporating one type of design structure, i.e., function structures.

With mobile-based applications gaining rapid popularity, voice integration can enhance efforts at supporting “designing in motion” and collaboration in globally integrated product teams, where designs interact with human teams seamlessly. The results from this work pave the way for such a future. It was shown that functional information about parts, sub-assemblies and entire products can be readily captured and interactively discussed with a user. The results also indicate that closely related words may sometimes be mis-classified. The performance of speech recognition can be significantly improved with the help of neural network-based learning tools that can be implemented within such a software package. Additionally, while currently part functions were readily chosen from a database of functions for the specific product, further work can involve creating this database dynamically with new users adding or removing new functions within the PLM system, as the need arises. Furthermore, maintenance issues and other product related data can also be embedded as voice in the CAD models and queried as needed. Additional layers of intelligence built within futuristic systems can also involve the software identifying which user is talking to it in a live conference meeting and responding to the user as per his role in the product development process.

VI. CONCLUSIONS

A new framework for voice integration of functional information in CAD models has been discussed in this paper. The results indicate the feasibility of doing this in a systematic manner by using speech recognition and text-to-speech conversion simultaneously. A live connection between MATLAB and SolidWorks was combined with voice based tools to achieve this. The use of MATLAB to establish the live connection enables easy access of variables in the CAD model that can then be used for interacting with the user. This work should help provide guidelines for next generation PLM tools that integrate voice seamlessly, especially within applications for hand-held devices such as cellphones and tablets, and set up a two-way active learning environment helping the user learn from the CAD model and also learning about the user.

ACKNOWLEDGMENT

The authors would like to thank the Engineering and Physical Sciences Research Council (EPSRC) for grant number EP/N005694/1 which supported this research.

REFERENCES

[1] S. Xue, X. Kou, S. Tan, Natural voice-enabled CAD: modeling via natural discourse, Computer-Aided Design and Applications, vol. 6, 2009, pp. 125-136.

[2] X. Kou, S. Xue, S. Tan, Knowledge-guided inference for voice-enabled CAD, Computer-Aided Design, vol. 42, 2010, pp. 545-557.

[3] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, S. Di Marca, Accelerometer-based gesture control for a design environment, Personal and Ubiquitous Computing, vol. 10, 2006, pp. 285-299.

[4] C.-C.P. Chu, T.H. Dani, R. Gadh, Multi-sensory user interface for a virtual-reality-based computeraided design system, Computer-Aided Design, vol. 29, 1997, pp. 709-725.

[5] X. Liu, G. Dodds, J. McCartney, B. Hinds, Manipulation of CAD surface models with haptics based on shape control functions, Computer-aided design, vol. 37, 2005, pp. 1447-1458.

[6] N. Diniz, C. Branco, An approach to 3D digital design free hand form generation, in: Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on, IEEE, 2004, pp. 239-244.

[7] S. Gao, H. Wan, Q. Peng, An approach to solid modeling in a semi-immersive virtual environment, Computers & Graphics, vol. 24, 2000, 191-202.

[8] M. Weyrich, P. Drews, An interactive environment for virtual manufacturing: the virtual workbench, Computers in industry, vol. 38, 1999, pp. 5-15.

[9] R.A. Bolt, E. Herranz, Two-handed gesture in multi-modal natural dialog, in: Proceedings of the 5th annual ACM symposium on User interface software and technology, ACM, 1992, pp. 7-14.

[10] X. Kou, S. Tan, Design by talking with computers, Computer-Aided Design and Applications, vol. 5, 2008, pp. 266-277.

[11] T.H. Dani, R. Gadh, Creation of concept shape designs via a virtual reality interface, Computer-Aided Design, vol. 29, 1997, pp. 555-563.

[12] G. Pahl, W. Beitz, Engineering design: a systematic approach, Springer Science & Business Media, 2013.

[13] B. Helms, K. Shea, F. Hoisl, A framework for computational design synthesis based on graph-grammars and function-behavior-structure, in: ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers, 2009, pp. 841-851.

[14] A.K. Behera, A. McKay, H.H. Chau, A. de Pennington, M.A. Robinson, Embedding Design Descriptions Using Lattice Structures: Technical Requirements, User Perspectives and Implementation, in: International Conference on Research into Design, Springer, 2017, pp. 557-566.

[15] S. Szykman, J.W. Racz, R.D. Sriram, The representation of function in computer-based design, ASME Design Engineering Technical Conferences, 11th International Conference on Design Theory and Methodology, 1999, New York.

[16] B. Eisenbart, K. Gericke, L.T. Blessing, T.C. McAloone, A DSM-based framework for integrated function modelling: concept, application and evaluation, Research in Engineering Design, vol. 28, 2017, 25-51.

[17] Y. Liu, A. Chakrabarty, T. Bligh, A strategy for functional design, in: Proceedings of the 12th International Conference in Engineering Design (ICED'99); Vol. 2, Technische Universitat Munchen, 1999, pp. 1187-1190.

[18] K.L. Wood, Development of a functional basis for design, Journal of Mechanical design, vol. 122, 2000, 359-370.

[19] A. Behera, A. McKay, H. Chau, M. Robinson, Embedding multiple design structures into design definitions: A case study of a collision avoidance robot, in: DS 84: Proceedings of the DESIGN 2016 14th International Design Conference, 2016. Dubrovnik, Croatia.

[20] B. Eisenbart, K. Gericke, L.T. Blessing, Taking a look at the utilisation of function models in interdisciplinary design: insights from ten engineering companies, Research in Engineering Design, 2016, pp. 1-33.

[21] M. Brookes, VOICEBOX: Speech Processing Toolbox for MATLAB, in, 1996 http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html, Last accessed 16th June, 2017.

[22] Finite Business Inc, CADLab - Toolbox that enables live connection between MATLAB and SolidWorks, 2010, https://uk.mathworks.com/products/connections/product_detail/product_66411.html, Last accessed 16th June, 2017.

[23] J. Gudnason, M. Brookes, Voice source cepstrum coefficients for speaker identification, in: Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, IEEE, 2008, pp. 4821-4824.

[24] D.M. Brookes, SAPISYNTH text-to-speech synthesize of text string or matrix, 2003, http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/doc/voicebox/sapisynth.html, Last accessed 16th June, 2017.

[25] S. Deng, Text-to-speech, 2007, https://uk.mathworks.com/matlabcentral/fileexchange/18091-text-to-speech, Last accessed 16th June, 2017.