Spoken Interaction with Broadcast Debates

Rolando Medellin, Chris Reed and Vicki Hanson
School of Computing, University of Dundee

Abstract.
The constant emergence and change of technologies in the form of digital products and services can cause certain groups of the population to feel excluded, older adults represent one such group. We investigate how to combine applied research on computational models of argument and human-centric computing to impact the way in which older adults interact with broadcast debates. This paper describes a technology application that uses a speech recognition interface to interact with broadcast debates. The application classifies spoken utterances and creates positive or negative "votes" related arguments from a debate. We describe a user study where older adults interact with a debate using the application. Our results indicate that the use of speech recognition, extra information provided to users, and feedback on their interaction, plays an important role in their engagement with the debate.

Keywords. Argument Web, Speech Recognition, Sentiment Analysis.

1. Introduction

Older adults represent a group in society that faces particular challenges when adapting to new technologies [4]. Audiences composed of such group, could be prevented from interacting with broadcast debates because of technological barriers imposed by current methods available to interact and participate. As broadcasters and governments increasingly use social media for communicating with their audience, it is worrying to note that social media use amongst older adults is still at very low levels. This is a group that is at risk of being systematically and increasingly excluded from participation. By allowing them to provide their opinion about topics that interest them, we want to provide them with a sense of engagement and, as well, a sense of empowerment to make them feel their voice is being heard. Older adults are a segment of the population that is not typically characterised by engagement with technology, with just over 5 in every 10 people aged 65-74 having used the internet in the UK and only 3 in every 10 over the age of 75 [9]. [1] These numbers however, are dropping, in the UK the Office for National Statistics reports that there was a decrease of 23,000 adults (0.3%) who had never used the Internet since Q1 2013, and a decrease of 748,000 adults (10%) compared with a year earlier (Q2 2012)[9]. Similarly, in the US, as of April 2012, only 53% of American adults aged 65

[1] For the purpose of this paper, we consider an *older adult* as an adult over 60 years old with no regard to their cognitive or physical conditions.

and older use the internet or email [17]. Older adults are still less likely than all other age groups to use the internet but data from 2012 represent the first time that half of seniors are going online after several years of very little growth among this group. Even more strikingly, this demographic makes little use of social media, in the UK only 13% used social networks like Facebook in 2013 [9].

In contrast, a large majority of the population have access to radio and/or TV [8]. The audiences of these programmes can participate in several ways mainly via email, phone and social networks. With the widespread use of social media platforms (e.g., Twitter and Facebook), the audience can participate more actively in debates but a generation of listeners are effectively being excluded from these interactions. Our aim is to provide a simple way to interact with broadcast debates by using speech recognition software that identifies verbalised interactions through their phone line. We discuss techniques to identify "sentiments" from speech in order to generate content related to arguments in AIFdb [5]. We use the term "sentiment" in reference to the predictive judgements about a topic. We aim to benefit from the inner desire in all of us to"shout back" when we agree or disagree with opinions expressed on the radio or TV and create content related to structured analyses in the Argument Web [12].

2. Interaction with Broadcast Debates

Current methods to obtain audience reactions from broadcast debates in the UK are carried out mainly via online forums, email, phone and just recently, social networks. [2] Internet forums or message boards, provide a way to interact with debates but these are limited since their structure do not provide an intuitive interface to represent debates and interact with them. The *BBC Message Boards* for example, provide access to discussions that can contain several hundred posts and makes difficult tasks such as browsing and analysing opinions. Furthermore, debates cannot be represented chronologically and the grouping of topics makes it difficult to refer to specific arguments. Users therefore can get lost easily in forums and lose their interest to participate.

Social media platforms can provide access to a large network of connected users to participate in broadcast debates due to their immediacy. As an example, the idea of Twitter as an "opinion thermometer" has been widely studied in the context of political elections e.g., [14,15]. As an example of engagement and participation, the 2012 presidential debates in the United States attracted 59 million viewers [3] and around 7.2 million *tweets* [4]. When people use social media to interact with live broadcast debates, they react to different arguments at different times making it difficult to classify specific reactions as the debate advances. The advantages in terms of immediacy provided by social media has not been integrated into solutions that benefit the audience (by getting feedback from their interactions) and the producers of programmes (by getting meaningful data to

[2]The BBC broadcasts eight debate programmes (March 2014): Question Time (TV, BBC1), Free Speech (TV, BBC3), The Big Questions (TV, BBC1), Sunday Morning Live (TV, BBC1), The World Debate (BBC News), Brian Taylor's Big Debate (BBC Radio Scotland), The *Moral Maze* (BBC Radio 4) and The Intelligence Square Debate (BBC World News).

[3]http://www.nielsen.com/us/en/newswire/2012/final-presidential-debate-draws-59-2-million-viewers.html [June 2014]

[4]http://www.digitaltrends.com/social-media/the-internets-reaction-to-last-nights-presidential-debates [June 2014]

improve their format and/or content). Engagement in social media is important since it has the potential to foster groups of older adults to feel included. But social media is not available to all audiences and certain groups can feel excluded if social media is the only way to do it. We identified two main issues in our task to encourage older adults that like to participate in broadcast debates to do it. First, the entry level to participate has to be extremely simple and the content from the debate has to be available so that we can obtain meaningful data from the interactions and give feedback to the participants.

3. Capturing Sentiments from Speech

Typically, verbal communication conveys more explicit information than writing, which is associated with a more formal and structured style. The use of spoken utterances to agree or disagree is more evident since the intonation and volume of the speaker can give a more clear indication of a person's position towards an argument. Our proposed solution to help older adults to interact with a debate is the following: (1) capture verbalised interactions using speech recognition, (2) classify automatically the sentiment expressed, (3) create argument-related content related to the analysed debate in AIFdb, and (4) present to the user feedback on their interactions and the interactions of others. The technology application is comprised of three main modules (see Figure 1):

- The ARGPlayer which synchronise: audio from a debate program, its transcript and the related analysis (i.e., the segmented arguments) in AIFdb.
- The Speech Recognition and Sentiment Analysis (SRSA) module captures user verbalised opinion and classifies the sentiments expressed at the time at which they are verbalised. The term *Sentiment Analysis* refers to the automatic analysis and classification of opinions and subjectivity in texts [11] and has been applied in the context of analysing market opinions [2], tracking online discussions [13] and Natural Language Processing techniques [7].
- The matching module, that links at runtime the corresponding argument in AIFdb with the spoken participation and creates AIF nodes related to debate arguments.

We used SPHINX 4 [16], an open source Speech Recognition API to recognise utterances. As for the acoustic model and the phonetic dictionary, we used the ones provided with the API distribution. We defined a grammar of words and phrases linked to a statistical Language Model from a previous study in [6]. The grammar is composed of classified phrases that suggest agreement or disagreement. To identify the sentiment, we compare the recognition from the speech engine (which is a mixture of speech recognition and keyword spotting techniques) with keywords in our grammar. From each spoken participation we identify either an exact phrase defined in our grammar or specific keywords from unrecognised phrases which we then parse searching for keywords and phrases to decide if the participation conveys a positive or a negative sentiment. Finally, the Synchronisation module relates the user interaction in the for of an AIF node with nodes from the analysed debate. To do so, we synchronise the start of the audio with the arguments from the debate in AIFdb, and each time an interaction is recognised, we generate AIF nodes related to the nearest previous argument in the time line of the debate.

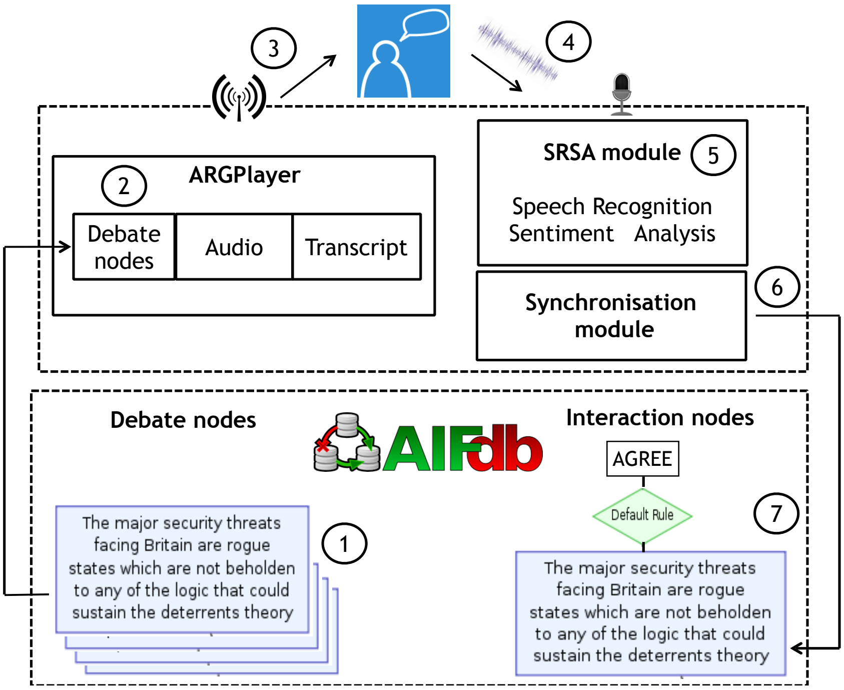

Figure 1. Application architecture. 1) The debate analysis in AIFdb, 2) ARGPlayer synchronizes AIF-nodes nodes with audio, 3) Participant listening to the Moral Maze program, 4) Spoken participation, 5) SRSA module classifies utterance, 6) Synchronization module relates current AIF-node with participation, 7) Participation node represented in AIFdb.

3.1. Representing User Interactions in AIFdb.

The user interactions obtained from the speech recognition module do not strictly represent arguments, nevertheless, we create nodes that represent the sentiment of the user related to a real argument. An argument in AIFdb is represented as an Information node (I-node) which is associated with a Locution node (L-node) that specifies the person that issues the propositional content in the I-node. Each participation creates AIF nodes that represent the sentiment of the participant in terms of a Rule Application (RA) node, (that denotes the application of a scheme, an agreement for our purposes) or a Conflict Application node (CA) (that denotes disagreement). To create these relations we follow Inference Anchoring Theory (IAT) [1] that allows the representation of inference and dialogical structures in the same model. IAT shows how the illocutionary force of *arguing* in dialogues can be used to anchor inferences between propositions. For each participation then, we create six AIF nodes (see Figure 2) that "anchor" the participant's sentiment to an AIF argument as follows: an I-node with its corresponding L-node, a RA or CA node according to the result of sentiment analysis, and a Transition node (TA) that relates the locutions. In our case the users interactions are represented as *arguing* relations to propositional content from the debate.

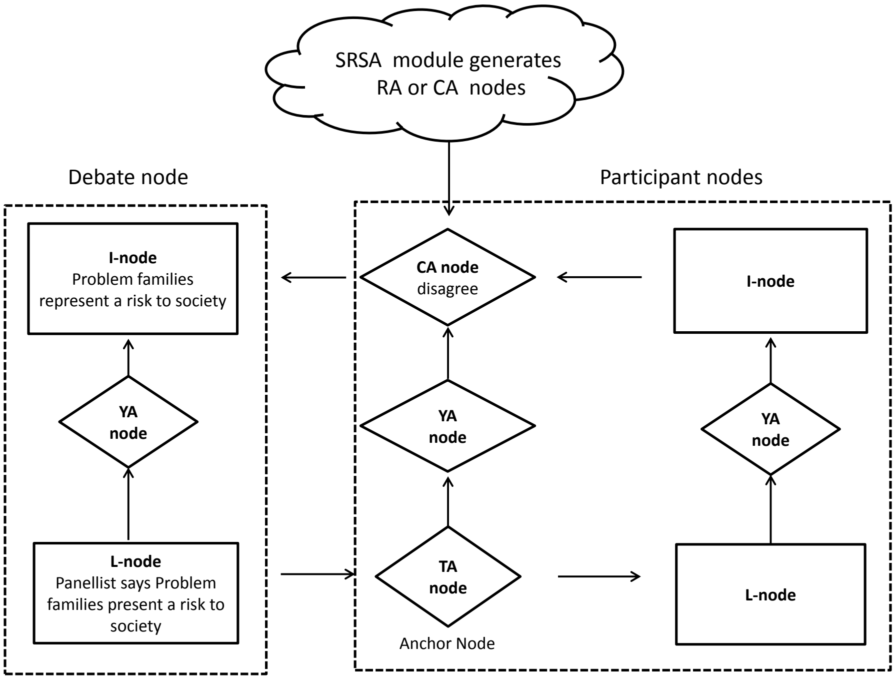

Figure 2. Anchoring speech interactions to arguments in AIFdb. The TA node is the Anchor node for the YA node which has as its content the RA or CA node.

4. User Study and Results

To test our application, we set up a study in which older adults interacted with *The Moral Maze* [5] using the application. We invited seven older adults that enjoy regularly to listen to broadcast debates to test the application. Five users participated individually and two in a pair. The participants listened for 20 minutes to a Moral Maze extract and interacted freely with it agreeing and disagreeing when they wanted. As a motivator we provided them with graphics or information related to the arguments listened in the form of agreement or disagreement rates from a previous survey. To measure the participants' level of engagement, we gave them questionnaires before and after the study. The objective of the study was to discover to what extent the application, information presented to them throughout the program and the final feedback on their interactions motivated them to interact and engage further with the debate.

To obtain the sentiment recognition accuracy, the researcher registered the sentiment from participants independently from the application recognition. We compared both recognitions and obtained an accuracy percentage for agreement and disagreement per user. Table 1 presents details on the recognition and accuracy for each participant. Column 2 presents the number of interactions per participant, i.e., the number of times the application recognised phrases separated by silences including those where no sentiment was recognised. Column 3 presents the number of agree/disagree recognitions made by

[5]The "*The Moral Maze*" is a BBC Radio 4 debate programme. http://www.bbc.co.uk/shows/moralmaze

the SRSA module; columns 5 -7 present details on the accuracy of the sentiment analysis recognition.

Table 1. Speech Recognition and Sentiment Analysis accuracy for a 20 minute *Moral Maze* extract.

Participant	# interactions	# rec.	% rec.	detail rec.	correct rec.	% correct
P1	82	53	64%	D: 49 A: 4	D: 17 A: 3	34 % 75 %
P2	138	53	38%	D: 47 A: 6	D:16 A: 4	34 % 66 %
P3	267	84	31%	D: 69 A :15	D:26 A: 12	26 % 80 %
P4	247	115	46%	D: 69 A: 46	D:23 A: 26	37 % 56 %
P5	222	72	38%	D: 41 A: 31	D:18 A: 20	43 % 64 %
P6, P7	286	170	32%	D : 83 A: 33	D:32 A: 26	38 % 78 %
avg	**180**	**78**	**43%**	**D: 59** **A : 19**	**D: 22** **A: 15**	**37.2 %** **78.9 %**

The questionnaires reveal that all participants wanted to know about what other people think, even if they did not have any interest in participating themselves. This is an interesting finding because the application and the information obtained from it can be used as a motivation to help drive engagement with the debate and ultimately with social media. We can use the data about how other people interacted with the debate, to compare how one listener's views compare with everyone else's as a feature to motivate the engagement and use of technology.

Regarding their interest to use their phone line to interact with debates, half of the participants expressed some interest after being shown how their verbalised opinions were related to the debate which suggest a positive impact on their perception about interaction with debates. The speech recognition and sentiment analysis modules perform reasonably well given the noisy environments in which they work. The recognition accuracy for disagreement was low with an average of 37% correct recognitions. The interactions when participants disagreed were in the form of several long phrases, making it harder for the SRSA module to classify the interaction. The fact that their interactions were in the form of long phrases reduced the accuracy of the speech recognition algorithm which was based on a grammar of short phrases. The keyword spotting module then needs to be extended and tested accordingly to allow identification in long phrases. The agreement expressions were in general shorter and the algorithm performed better, avg. 78.9 % identifying these interactions.

It would be interesting to explore the accuracy with keyword spotting open source solutions, since our aim is not to generate a speech to text application but an application that identifies specific keywords that can lead to identify arguments.

5. Related and Future Work

Recent work on argumentation has been focused on combining sentiment analysis and argument theory to identify arguments and opinions, e.g., [3,10] . In contrast, our aim was to identify sentiments from speech automatically while users listen to a debate. These sentiments can be used later as the start point to identify arguments and opinions.

Given the multidisciplinary nature of the project there are several research paths to follow. Regarding the argumentation content generated, we found that the interactions did not always refer to the arguments presented in the debate when we synchronised the times. Participants express their opinion not only at the time when they hear something they agree or disagree with but also at the end of a witness participation or even at the end of the program, making it harder to link the interaction to specific arguments using just the time at which it is being verbalised. Therefore, we need to create content related to sets of arguments or persons rather than to specific arguments to be more precise in capturing the general sentiment of the user.

We are also interested in extracting more information from the user interaction such as keywords or phrases that suggest participants are arguing with specific schemes. Our plan is to take this pilot study and extend it using a larger group of participants and experiment with different scenarios (e.g., social sessions after and before the study, immediate feedback embedded in the original audio, etc.) to better evaluate the benefits of the application. As for the sentiment analysis accuracy it would be interesting to integrate existing pre-labelled grammars to the application and use machine learning techniques to train and evaluate our approach.

6. Conclusions

We described a technology application based on argument representation and speech recognition that supports interaction with debates. The application captures utterances from participants and generates AIF-content related to an existing analysis that represents the debate on AIFdb. Furthermore, we described a pilot study where users experimented with the application and received feedback on their interactions and the interactions of other users. Preliminary results suggest that participants interested in interacting with broadcasts debates could benefit from interacting using a speech recognition tool since they feel their opinion is being taken into account. While there remains work to do in the speech recognition an sentiment analysis module, we set the basis to explore ways in which older users can benefit from participating in broadcast debates. By doing so we hope not only to equip an undeserved part of the audience with the ability to have the voice heard, but also in the longer term to provide a driver for engagement with social media for this user group.

7. Acknowledgements

This research was supported by RCUK grant EP/G066019/1 “RCUK Hub: Inclusion in the Digital Economy.” Rolando Medellin is grateful for support from the Mexican National Council of Science and Technology (CONACYT). The authors would like to thank the anonymous reviewers for their feedback and Christine Morgan, executive producer of the BBC’s *Moral Maze*, for her support.

References

[1] K. Budzynska and C. Reed. Speech Acts of Argumentation: Inference Anchors and Peripheral Cues in Dialogue. In *Computational Models of Natural Argument*, 2011.

[2] S. R. Das and M. Y. Chen. Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web. *Management Science*, 53(9):1375–1388, Sept. 2007.

[3] K. Grosse, C. I. Chesñevar, and A. G. Maguitman. An argument-based approach to mining opinions from twitter. In *AT*, pages 408–422, 2012.

[4] V. Hanson. Influencing technology adoption by older adults. *Interacting with Computers*, 22(6):502–509, Nov. 2010.

[5] J. Lawrence, F. Bex, C. Reed, and M. Snaith. AIFdb: Infrastructure for the Argument Web. In *Computational Models of Argument COMMA 2012*, pages 515–516, 2012.

[6] R. Medellin-Gasque, C. Reed, and V. L. Hanson. Guidelines to support older adults interaction with broadcast debates. *Manuscript under Review*.

[7] T. Nasukawa and J. Yi. Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In *Proceedings of the Conference on Knowledge Capture (K-CAP)*, 2003.

[8] Ofcom. Media Literacy Audit: Report on media literacy amongst older people. (April), 2006.

[9] Office for National Statistics. Statistical Bulletin Internet Access Quarterly Update, Q2 2013. (August), 2013.

[10] A. Pak and P. Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the 7th Conference on Language Resources and Evaluation*. European Language Resources Association, 2010.

[11] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and trends in Information Retrieval*, 2(1-2):1–135, 2008.

[12] I. Rahwan, F. Zablith, and C. Reed. Laying the foundations for a World Wide Argument Web. *Artificial Intelligence*, 171(10-15):897–921, July 2007.

[13] R. Tong. An Operational System for Detecting and Tracking Opinions in On-line Discussions. In *Working Notes of the SIGIR Workshop on Operational Text Classification*, pages 1–6, New Orleans, Louisianna, 2001.

[14] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting Elections with Twitter : What 140 Characters Reveal about Political Sentiment. In *The International AAAI Conference on Weblogs and Social Media*, pages 178–185, 2010.

[15] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin. Canal9: A database of political debates for analysis of social interactions. *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pages 1–4, Sept. 2009.

[16] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel. Sphinx-4: A Flexible Open Source Framework for Speech Recognition. Technical report, Mountain View, CA, USA, 2004.

[17] K. Zickuhr and M. Madden. Older adults and internet use. Technical report, Pew Research Center Internet and American Life Project, 2012.