

TeMoCo: A Visualization tool for Temporal Analysis of Multi-Party Dialogues in Clinical Settings

Shane Sheehan*, Pierre Albert*, Masood Masoodian[†] and Saturnino Luz*

*Usher Institute, Medical School, The University of Edinburgh, UK

Email: {Shane.Sheehan,Pierre.Albert,S.luz}@ed.ac.uk

[†]School of Arts, Design and Architecture, Aalto University, Finland

Email: masood.masoodian@aalto.fi

Abstract—We present a tool for visualization of transcripts of multi-party dialogues, with application to the analysis of communication in medical teamwork. The visualization is based on a “temporal mosaic” metaphor, which provides a temporal overview of dialogues and supports the tasks of transcript browsing and information access, by segmenting the dialogue and laying out the keywords of the different segments on interactive visual “tiles”. The tool has been tested on a corpus of transcribed dialogues among the members of a (simulated) critical care team. An analytical evaluation is presented which demonstrates the potential uses of the tool in an educational setting and highlights areas for improvements.

Keywords—Medical team communication; temporal visualization; temporal mosaics; speech visualization.

I. INTRODUCTION

Effective verbal communication is crucial to the success of clinical encounters, including clinician-patient consultations, multidisciplinary medical team meetings, accident and emergency contexts, among others. Analysis of communication in such contexts is important for care quality assessment, individual appraisal, assessment of interventions, as well as training and education. However, this kind of analysis tends to be very time-consuming, requiring substantial input from healthcare research experts. While frameworks such as the widely used Roter Interaction Analysis System (RIAS) [1] have helped standardize and guide such work, analyzing medical communication at scale remains a challenge.

While recent advances in speech and language processing technologies promise to facilitate the job of healthcare communication analysts [2], visual tools are still needed to harness the power of these technologies, without requiring analysts to understand their underlying complexity.

Here, we introduce a new visualization, called *TeMoCo*, which aims to support temporal analysis of conversations. We have developed an interactive prototype tool based on this visualization which is designed for use in clinical settings. We present this prototype, illustrate its use with a case scenario of analysis using a selected corpus of multi-party dialogue of conversations in an A&E unit, and perform a cognitive walk-through to evaluate the prototype.

II. COMMUNICATION IN CLINICAL SETTINGS

Clinical conversations play a major part in medical communication, and extensive literature exists on the topic. Medical communication is a complex process, with both biomedical objectives (e.g. establishing a diagnosis, curing the patient) and humanistic objectives (e.g. mutuality of the relationship, effective communication). The communication is impacted by widely different aspects related to socio-demographic, cultural, and even personality aspects, and will vary with diseases-related characteristics, such as the stage of the illness and patient expectations [3]. In medical team settings, communication takes place while team members are cooperating toward a common goal. For example, in A&E the common goal is saving the patient while conducting a number of tasks under complex constraints, including time pressure, information overload, ambiguous situations, and the risk of severe consequences in case of error. The impact of poor communication in such settings is evident.

Teaching and training for good communication skills are therefore necessary. In medical education, training happens at different stages of the professional life of doctors and nurses. It usually includes simulated interventions, where technical and non-technical skills are assessed. To evaluate medical communication training sessions, and provide feedback, the health community has looked at systematic analysis and problem-solving approaches developed in other life critical domains (e.g. crew resource management from aviation) and have implemented similar solutions in clinical settings [4].

Dedicated frameworks have been developed for the assessment of communication – often referred as non-technical skills – each of which assess different sets of skills. For instance, the Observational Teamwork Assessment of Surgery (OTAS) [5] assesses clinical and technical skills, and also interpersonal skills and behaviours. The Anesthesiologists Non-Technical Skills (ANTS) [6] assesses four different sets of skills: task management, team working, situation awareness, and decision making. Similarly, the Communication And Teamwork Skills (CATS) [7] assesses four sets of skills: situation awareness, coordination, communication,

and cooperation. The above mentioned RIAS framework [1] has also been used in assessing communication skills in these settings.

Despite these efforts, overall, there is a lack of consensus regarding the evaluation of clinical team communication [8], and so far, there are no globally accepted theoretical models for assessment of team performance. However, general studies of team performance and studies specific to healthcare have identified some necessary skills. These studies rely on the observation and analysis of certain conversational behaviours during training encounters.

In this context, the use of tools to extract and visualize conversations can support the temporal analysis of team communication. A visualization tool could provide the analyst with a simple and natural way to navigate conversation sessions and to search for different interactional aspects related to the monitored skills – either punctual (e.g. verbalization of plans and changes, requests for help, use of key phrases) or spanning the whole interaction (e.g. acknowledgement of the concerns of others, closed-loop communication, updates).

A. Related work on visual tools for clinical communication

The most common method used for medical communication training is the video recording of a session followed by an after-action-review. The review is performed either by a professional, or provided to the students – e.g. to write self-reflective structured assessment of their performance. Visualization tools exist to help with the analysis of sessions.

The Lab-in-a-box system [9] uses sensors (3D camera, eye tracking, computer activity) to track the clinician’s workflow during a medical consultation. The collected data are presented as events along a timeline representing the consultation. Simple events (key strokes and mouse clicks) are presented directly, and visual attention toward the computer is displayed as blocks. The selection of a block opens a picture showing the corresponding gaze direction track.

EQClinic [10], a fully-fledged system for training of health professionals, records and analyses online sessions with simulated patients. Live feedback is provided by the assessor in the form of comments with positive or negative valence. Post-interaction tools includes manual assessment (forms) and automated analysis of non-verbal communication (turn patterns, prosody, visual cues).

While these systems support complex analytic tasks, the display of single features separately without context is difficult to interpret by non-experts. To facilitate interpretation, specific visualizations of the content of the interactions need to be developed. Addressing this issue, Angus et al. [11] provided a visual representation of content to track conceptual recurrence in the conversation structure of medical consultations. Our approach also aims to address the issue of providing a temporal structure to dialogue content, but we employ a different visual representation that scales

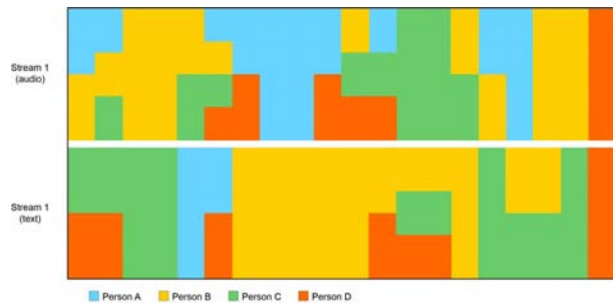


Figure 1. A sketch of temporal mosaic visualization, adapted from [12].

to dialogues with more than two participants (multi-party dialogues), as explained below.

III. *TeMoCo* VISUALIZATION

We have designed the *TeMoCo* (Temporal Conversation Mosaics) visualization to better support visual analysis of conversations. It uses the temporal mosaics visualization [12] as its basis. The original temporal mosaics visualization (see Figure 1) represents the individual time-based data streams separately as synchronized rows of visualizations.

In the case of Figure 1, the top row shows the audio conversations between 4 people, while the bottom row shows their contributions to a text document. Within each row, the temporal mosaics visualization allocates the vertical space equally between the number of participants active in each time-slice (i.e. the horizontal space) – with the resulting sum of individually coloured rectangular shapes showing the contributions of each participant across time. A temporal mosaic visualization is, therefore, used to represent temporal contribution patterns rather than the content of individual contributions. However, when used as an interactive visualization [13], each rectangular segment of a temporal mosaics visualization can be linked to the corresponding part of the data stream it represents – thus supporting access to media content, both temporally as well as contextually.

In the case of analysis of audio recorded conversations, we are only dealing with a single data stream. As such, *TeMoCo* can utilize the visualization space to represent a single data stream using the convention of dividing the vertical space equally between the active conversation participants for each time-slice – similar to the top row of Figure 1. While in an interactive version each mosaic segment can be linked to its corresponding audio recording, *TeMoCo* uses the visual space of each segment to also superimpose a textual summary of the transcript of the corresponding audio speech, making it more useful even in a static mode. This textual summary can take a number of forms, depending on the application area for which the visualization is used. Here, we have chosen to provide a list of keywords from each speech segment, ranked according to their occurrences. Other options could include a word-cloud of keywords for

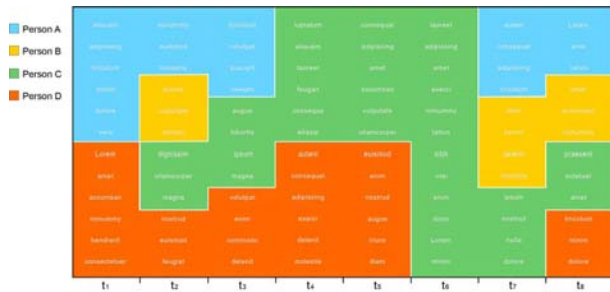


Figure 2. A sketch of *TeMoCo* visualization.

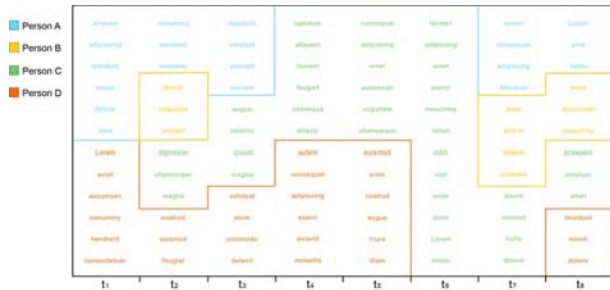


Figure 3. An alternative version of *TeMoCo* with coloured keywords.

each segment (e.g. in a manner similar to Wordle [14]), a representative sentence (e.g. first sentence), etc. Figure 2 provides a sketch of the *TeMoCo* visualization.

Even though *TeMoCo* is visually similar to a temporal mosaic visualization, the additional visual encodings must be carefully considered. The first issue to consider is the visual contrast between the text and the background mosaics. Although an increased contrast would make the text more readable, it would also cause visual distraction – thus reducing visual detection of the background mosaic patterns. As mentioned, detection of these mosaic patterns is an important aspect of the original temporal mosaics visualization, allowing the user to easily view contributions of each of the conversation participants across time, to detect, for instance, any imbalance in levels of contribution, dominance of one participant, and so on. Therefore, although in Figure 2 we use white text on colour mosaics with only hue variations between their colours, ultimately such variations need to be adjusted to suit the static or interactive uses of the visualization. For example, Figure 3 shows an alternative version of *TeMoCo* which might be better for printing in static form.

Another issue to consider in *TeMoCo* visualization is the choice of the number of words for each mosaic segment, as well as the size and style of typefaces used to show the selected words. Further to considering the issue of contrast discussed above, these variations are dependant on the visual and temporal length of each time-slice. Increasing the length of time-slices visually allows for better accommodating



Figure 4. The *TeMoCo* prototype, with the visualization on the left, and transcripts pane on the right.

longer keywords and/or making their typeface size bigger – thus increasing readability. However, increasing the visual length of time-slices may require increasing their temporal length as well. This in turn has its own consequences. For instance, longer temporal time-slices would have longer transcripts to be represented (e.g. requiring more keywords to be selected). Furthermore, if the time-slices are too long, then they may end up including every conversation participant in each slice, and as such, reduce visual effectiveness of mosaic patterns. Once again, these issues are application dependant and must be considered for each use case.

A. Prototype

We have developed an interactive prototype tool which uses *TeMoCo* to visualize multi-party conversations, aiming at supporting the visualization of communication among medical team members. Figure 4 shows the interface of the *TeMoCo* prototype. As can be seen, the left-hand panel is the interactive visualization showing the temporal mosaic patterns of the conversation – along with the top keywords selected from each speaker turn – and the right-hand panel shows the transcript of the entire conversation session. In this conversation session there are five participants (Patient 1, Nurse 1, Doctor 1, Doctor 2, and Medical Registrar 1), who have been talking for 13 minutes and 30 seconds.

While the static view of *TeMoCo* is sufficient for seeing the patterns of conversation, and a summary of its main keyword points, the user can get a detail-on-demand view by clicking on a speaker turn mosaic on the visualization to access the relevant parts of the conversation on the transcript. Figure 5 shows a selected mosaic (in gray colour) on the left for participant D1, between 04:30 and 06:00. By selecting a mosaic, the prototype tool locates the start of the transcript text related to the selected time-slice (04:30–06:00), grays out the background of all the text for that time-slice, and then highlights the segments of the transcript text for the chosen speaker during that time-slice using the colour assigned to that speaker (the orange colour for D1 in Figure 5, the blue colour for P1 in Figure 6).



Figure 5. The *TeMoCo* prototype, with a speaker turn selected on the visualization (grayed out mosaic on the left), and the relevant parts of the transcripts highlighted (orange background text on the right).



Figure 6. The *TeMoCo* prototype, with another speaker turn selected on the visualization (grayed out mosaic on the left), and the relevant parts of the transcripts highlighted (blue background text on the right).

B. Implementation

Figure 7 shows the architecture of the *TeMoCo* prototype which has been implemented as a single-page web application using the *D3.js* framework [15]. The current system creates the visualization using a transcript file made available to it on the server. The transcript text is time-stamped and tagged with the labels of the conversation participants.

The system starts by pre-processing the transcript text to create two data streams. The first stream generates a data source containing relevant keywords, in which the keywords

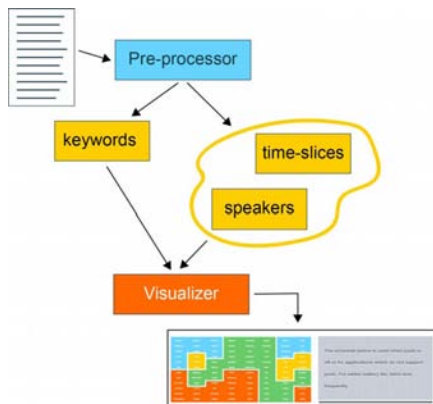


Figure 7. Architecture of the *TeMoCo* prototype.

are selected for each time-slice and participant combination. Keyword salience is dependant on context and use case – measures such as word frequency or frequency in a domain specific reference corpus are an obvious starting point. In our tests we found raw frequency to be uninformative, and subsequently decided on the manual selection of seemingly salient words, this simulates the word selections that could be achieved automatically using a medical reference corpus. Depending on the corpus and use case, any statistical measure of word salience or keyness could be injected to produce the keywords for a speaker in a time-slice.

The second stream of data is generated by extracting the time-slice and speaker information. This information is then used for tagging the input transcript with HTML attributes. This enables dynamic manipulation of the raw transcript as a part of the system interface.

Once the two data streams have been processed, the system constructs the temporal mosaics of the *TeMoCo* visualization from the time-slices, speakers and keywords information. The visualization and transcript panels are then positioned in the same web page. Both views are linked via the data, allowing interactions between the two. Selection of a time-slice mosaic scrolls the transcript to the corresponding time-slice, as describe above.

IV. EVALUATION

We have conducted an initial evaluation of *TeMoCo* using an existing corpus of transcribed medical conversations. We employed the cognitive walk-through analytical evaluation methodology to assess the use of *TeMoCo* on these data, in a medical education scenario.

A. Test conversations data set

To test our visualization, we selected a corpus of multi-party dialogues recorded in a hospital in Ireland, as part of the INCA (Interaction Analytics for Automatic Assessment of Communication Quality in Primary Care) project. The corpus was created for the development of tools for automatic analysis of verbal and non-verbal communication, to assess communication quality in different contexts of medical interaction. The corpus consists of simulation-based team training sessions for health professionals intervening in medical emergencies (e.g. accident and emergency services).

Each session follows a scenario with a specific medical problem selected by teaching staff. At the start of the training, the simulated patient – a dummy on a bed played by an actor outside the room – is showing rapid signs of health deterioration. The team must jointly establish a diagnosis and provide relevant care. Vital signs of the simulated patient are displayed on a patient monitoring equipment by the bed. Each recording features a nurse and two doctors. The nurse is present from the beginning and calls a doctor after detecting the abnormality. As the problem get more serious, a second doctor is called, If specific difficulties or questions

Total	Duration		Participants	Utterances
	Session	Turn		
207min	14min47s (11'20 - 23'01)	01.87s (1.1 - 3.9)	5.6 (5 - 8)	450 (296 - 680)

Table I
A SUMMARY OF THE CHARACTERISTICS OF THE INCA MEDICAL CONVERSATIONS CORPUS.

arise, the medical team could call specialist registrars (e.g. anaesthetic, orthopedics, etc.) on a telephone. A third doctor is sometimes present, as well as a second medical registrar.

A total of 14 training sessions have been recorded, segmented and transcribed. An overview of the main statistics of the data set used in our testing is provided in Table I.

B. Cognitive walk-through

Cognitive walk-through requires setting the objectives and task for the user of the system, and walk through each to assess the usability and capacity of the system to fit its role.

1) *Persona*: A persona is a prototypical user whose knowledge and behaviour are representative of the target users of the system. The following personae were identified:

Trainer: Dr Grey

The trainer is a professor and a doctor, expert in the field of medical education. She is 40 years old, has trained students for 5 years and has already established routines for each pedagogic goal. She has average competency in computing, and will use the software as a tool to improve the impact of her feedback. She is not interested in the underlying technology, and the tool interface must be easy to understand and use for her purposes. During the training she will observe the session and take notes on paper regarding the different skills under evaluation. After the session, she will want to navigate the session and illustrate global and punctual aspects of the communication and certain events that happened during the session, either good or bad.

Learner: B. Baggins

The learner is a medical student, she has already studied for 4 years and finds it difficult to see the value of non-medical skills. Once she has completed the training session successfully, she needs to be given feedback on her own behaviour, as well as her role in the team. She needs to visualize directly the points made by the trainer.

2) *Individual actions*: We defined a set of criteria for evaluating the training interactions between the two personas participating in debriefing sessions. Each of the training interactions are then evaluated using these four criteria:

- 1) Is the effect of the user's action corresponding to the user's goal?
- 2) Is the action visible?
- 3) Is the action identifiable as being the correct one?
- 4) Is the feedback understandable?

Each task and comments on each of these criteria are presented below:

Access to the session and overview of the conversation and its main points. The trainer starts the *TeMoCo* prototype, and both personas look at the visualization, with no further action needed (see Figure 4). The trainer wants to see the general structure of the conversations and identify any global patterns (e.g. distribution of speech related to communication and cooperation).

- 1) the effect is immediate with the visualization and corresponding transcript visible.
- 2) yes.
- 3) if a single session is accessed, yes. If multiple sessions are available, a label with the ID/date/participants of the sessions would be needed to identify the correct session.
- 4) feedback is natural: the legend displays each participants labelled with a single colour, corresponding to the one used in the visualization. Temporal visualization of conversations is immediately visible. For each mosaic segment, the set of keywords provides an overview of the main items in the conversation. A visualization of the links between recurring terms – or semantically similar terms – would help materialize this recurrence.

Navigation through the session to select points of interest (e.g. new problems, new participant, etc.) The trainer uses the visualization and the shown keywords to select a specific segment where something of interest occurred. The use of task specific keywords (e.g. medical terms, requests, concerns) illustrates cooperative behaviour. The selected area is grayed out, and the participant's utterances of the corresponding time-slice are displayed and highlighted using the participant's colour in the transcript window (see Figure 5). The learner can see the focus of the current point being discussed.

- 1) yes.
- 2) yes.
- 3) the trainer will need to search the keywords across participant's utterances that were produced in the selected time-slice. Highlighting salient keywords in the transcript would help to contextualize them.
- 4) yes.

Illustration of a specific exchange over a few conversation turns by going through the details of the conversation. The trainer is looking at the visualization to pick up keywords, and scrolls the transcript window to use the corresponding detail of the conversation and switches between participants to visualize their turns. The trainer will use the mosaic segments of visualization and the transcript window to search for events marked in her notes (see figures 5 and 6).

- 1) yes.
- 2) yes, the side bar and utterances of the selected participant scrolls.
- 3) yes, but part of the utterances of a time-slice do

not fit in the browser. However, utterances belonging to the selected time-slice are delimited by a gray background.

- 4) yes, the side bars are standard and are commonly used in most interfaces.

Identification of conversation patterns from the global interaction to local interactions (e.g. cooperation, coordination, turn-taking behaviour, etc.) The trainer and learner look at the visualization to see the global structure of the conversations (e.g. contributions of each participants, occurrences of keywords, etc.). Specific behaviour leading to patterns of interest (turn taking behaviour) is accessed through the transcript (see Figure 6)

- 1) Utterances of other speakers are grayed out, making it difficult to see the sequence of speakers. The use of faded colour would allow an easier interpretation while keeping the significance of the gray/user coloured duality.
- 2) yes.
- 3) yes, but the user will need to read and point her selected local points of interest.
- 4) Local points of interest that are not shown on the visualization require to scroll through the transcripts and may be difficult to find within long temporal visualization time-slices.

V. CONCLUSION

We presented a temporal text visualization tool to support the analysis and exploration of transcripts of medical team communication. Testing was carried out on data from simulated situations in a critical care (accident and emergency) setting, and the usefulness and usability of *TeMoCo* to support medical education was assessed. The mosaic-based design was found to be effective in providing a contextualized, temporal view of the conversation. Future work will focus on incorporating further textual structure, such as topics and conversational threads, to the visualization, and on testing the tool in other analysis tasks, such as assessment of patient-doctor communication.

VI. ACKNOWLEDGEMENT

This paper is supported by European Unions Horizon 2020 research and innovation programme under grant agreement No 825153, project EMBEDDIA (Cross-Lingual Embeddings for Less-Represented Languages in European News Media). The results of this paper reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains. Pierre Albert has been funded through the INCA project. We thank the INCA project members in Ireland for granting us access to the trainee data.

REFERENCES

- [1] D. Roter and S. Larson, "The roter interaction analysis system (RIAS): utility and flexibility for analysis of medical interactions," *Patient education and counseling*, vol. 46, no. 4, pp. 243–251, 2002.
- [2] P. Ryan, S. Luz, P. Albert, C. Vogel, C. Normand, and G. Elwyn, "Using artificial intelligence to assess clinicians' communication skills," *BMJ*, vol. 364, 2019.
- [3] L. M. L. Ong, J. C. J. M. de Haes, A. M. Hoos, and F. B. Lammes, "Doctor-patient communication: A review of the literature," *Social Science & Medicine*, vol. 40, no. 7, pp. 903–918, Apr. 1995.
- [4] D. M. Gaba, S. K. Howard, K. J. Fish, B. E. Smith, and Y. A. Sowb, "Simulation-Based Training in Anesthesia Crisis Resource Management (ACRM): A Decade of Experience," *Simulation & Gaming*, vol. 32, no. 2, pp. 175–193, Jun. 2001.
- [5] A. N. Healey, S. Undre, and C. A. Vincent, "Developing observational measures of performance in surgical teams," *BMJ Quality & Safety*, vol. 13, no. suppl 1, pp. i33–i40, 2004.
- [6] R. Flin and N. Maran, "Identifying and training non-technical skills for teams in acute medicine," *BMJ Quality & Safety*, vol. 13, no. suppl 1, pp. i80–i84, 2004.
- [7] A. Frankel, R. Gardner, L. Maynard, and A. Kelly, "Using the Communication and Teamwork Skills (CATS) Assessment to Measure Health Care Team Performance," *The Joint Commission Journal on Quality and Patient Safety*, vol. 33, pp. 549–558, Sep. 2007.
- [8] D. P. Baker, S. Gustafson, J. Beaubien, E. Salas, and P. Barach, "Medical teamwork and patient safety: the evidence-based relation," *AHRQ publication*, vol. 5, no. 53, pp. 1–64, 2005.
- [9] N. Weibel, S. Rick, C. Emmenegger, S. Ashfaq, A. Calvitti, and Z. Agha, "LAB-IN-A-BOX: semi-automatic tracking of activity in the medical office," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 317–334, 2015.
- [10] C. Liu, R. L. Lim, K. L. McCabe, S. Taylor, and R. A. Calvo, "A web-based telehealth training platform incorporating automated nonverbal behavior feedback for teaching communication skills to medical students: a randomized crossover study," *Journal of medical Internet research*, vol. 18, no. 9, 2016.
- [11] D. Angus, B. Watson, A. Smith, C. Gallois, and J. Wiles, "Visualising conversation structure across time: Insights into effective doctor-patient consultations," *PloS one*, vol. 7, no. 6, p. e38014, 2012.
- [12] S. Luz and M. Masoodian, "Visualisation of parallel data streams with temporal mosaics," in *Proceedings of the 11th International Conference Information Visualization*, ser. IV '07, July 2007, pp. 197–202.
- [13] —, "A model for meeting content storage and retrieval," in *Proceedings of the 11th International Multimedia Modelling Conference*, ser. MMM '05, Jan 2005, pp. 392–398.
- [14] F. B. Viegas, M. Wattenberg, and J. Feinberg, "Participatory visualization with wordle," *IEEE Transactions on Visualization and Computer Graphics*, vol. 15, no. 6, pp. 1137–1144, 2009.
- [15] M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301–2309, Dec. 2011.