Collective Speech

Andrew Konya*[1] and Aaron Slodov†[2]

[1]Liquid Crystal Institute, Kent State University
[2]Electrical Engineering and Computer Science, Case Western Reserve University

1 Introduction

The ability of humans to communicate their thoughts in words is a defining characteristic of our species' intelligence. From an early age conversation serves as a primary method for learning information, building relationships, and coordinating action [1]. For this reason, the ability to participate in conversation stands as one of the primary requirements for an agent to be considered intelligent [2]. While these conversational intelligence tests are typically applied to purely digital systems, an increasing number of algorithms treat humans as computing agents, and the line between human and artificial intelligence has blurred [3]. One might refer to a group of sufficiently connected people as a single intelligent agent, yet the ability for groups to participate in conversational tests of intelligence is just developing. Recent work has demonstrated the ability of a group of people to complete linguistic tasks such as labeling pictures [4] and answering questions [5] within a conversation. Even so, a clear model of collective speech, or what it means for a group of people to *speak as one* is still lacking in the literature.

This work presents a theoretical model to consider what it means for a group of people to speak as one, and uses it to develop a general model of collective speech. Due to the semantic ambiguity in what it means to "speak" our approach is to construct a simple model of speech for a single person then extend it to groups. A platform for achieving real-time collective speech based on this model is then introduced.

2 Speech

Webster's defines speech as: *the communication or expression of thoughts in spoken words.* For the purpose of this work we will broaden this definition to include words which are not spoken, but dictated in some other form. We will continue with the following definition, speech: *the communication or expression of thoughts in words.* In the same vein of reasoning we also extend the definitions of speech's semantic brethren 'speak' and 'say' to include all forms of verbal expression. A proper understanding of how the mind contains thought and produces speech is a topic of wide debate [6]. Thus, this section will construct a simple experiential model of how one might produce speech, rather than attempt a substantial description of human cognition and dictation.

At any moment the human mind contains a set of information [7] colloquially referred to as memory or thoughts. While this information can be considered as either continuous or discrete [8] we will adopt a discrete description, letting a state of mind be given by a set of thoughts $\mathbf{T} = \{t_1, t_2, ..\}$ collectively referred to as a *thinking.* In general a thought is any type of conceivable information, however some thoughts may be linguistic in nature and are therefore able to be spoken. Let this subset of speakable thoughts be denoted by $\mathbf{S} \subset \mathbf{T}$. Furthermore, any two thoughts may have varying degrees of conceptual agreement between them [9]. We represent this thought semantic distance (tSD) between thoughts t_i and t_j as $\sigma(t_i, t_j) \in [0, 1]$ where $\sigma(t_i, t_j) = 0$ if t_i and t_j are in perfect agreement and $\sigma(t_i, t_j) = 1$ if t_i and t_j are in perfect disagreement.

Assuming the thoughts which are spoken are those which are most consistent with the speakers current thinking, one should consider a thinking semantic distance (TSD) which is a measure of conceptual similarity between a single thought and a thinking. The simplest construction of the TSD is to consider it as the average of the tSD's between the thought under consideration and all thoughts contained within the thinking. Such a TSD between a thought t_i and a thinking comprised of M thoughts is written as:

$$\psi(t_i, \mathbf{T}) = \frac{1}{M} \sum_{j=1}^{M} \sigma(t_j, t_i) \tag{1}$$

However, a more general TSD construction, of which eq.1 is a special case can be formulated. Letting $P(\sigma|t_i, \mathbf{T})$ be the normalized distribution of tSD's (σ's) between thought t_i and thinking $\mathbf{T}$, the

*andrew@remesh.org
†aaron@remesh.org

TSD can bet written as:

$$\psi(t_i, \mathbf{T}) = \int_0^1 f(\sigma)P(\sigma|t_i, \mathbf{T})d\sigma \qquad (2)$$

Where the convoluting function $f(\sigma)$ is a single valued function for $\sigma \in [0, 1]$. Note, eq. 1 is recovered when $f(\sigma) = \sigma$. With this, we proceed under the assumption that one will express the thought, of those contained in $\mathbf{S}$, which is most consistent with their thinking, ie. has the smallest semantic distance. The resulting speech function,taking a thinking $\mathbf{T}$ as input and outputting a spoken thought, is then written as:

$$\gamma(\mathbf{T}) \equiv \underset{t\in\mathbf{S}}{\operatorname{argmin}}\, \psi(t, \mathbf{T}) \qquad (3)$$

Note that while this model appears to treat speech as singular and discrete, by considering eq. 3 as a description of speech for a discrete window of time a dynamic model of continuous speech can be constructed as well.

3 Collective Speech

Extending the aforementioned model of speech to a collective agent (CA) comprised of a group of people requires a construction of the thinking $\mathbf{T}$ for the CA. Letting $\mathbf{T}^k$ be the thinking of the k^{th} of N individuals comprising the CA, the CA thinking is trivially $\mathbf{T} = \{\mathbf{T}^1, \mathbf{T}^2, ..., \mathbf{T}^N\}$. Consequently, the speech function (eq .3) of an individual can be applied to a CA as well. Clearly, computing with thoughts contained in the human mind is not trivial, so one cannot (yet) digitally construct the thinking for the CA. Therefore, realization of collective speech must take advantage of the quantified nature of our model.

Computing the speech function requires 1) obtaining a set of speakable thoughts from the CA's thinking, and 2) computing the TSD's for each of those speakable thoughts with respect to the CA's thinking. The first of this is trivial, and can be achieved by allowing members of the CA to input linguistic thoughts by typing them into an input box, then collecting the inputs in a database. Obtaining a semantic distance between a text input and any single thought contained within a person's thinking is computationally unreasonable. Therefore, a reformulation of eq. 1 in terms of individuals TSD's is required. Letting t_i^k be the i^{th} thought of M^k total thoughts contained in the k^{th} individuals thinking, the TSD between thought t_i and CA thinking $\mathbf{T}$ can be written as:

$$\psi(t_i, \mathbf{T}) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{M^k}\sum_{j=1}^{M^k}\sigma(t_j^k, t_i) = \frac{1}{N}\sum_{k=1}^{N}\psi(t_i, \mathbf{T}^k) \qquad (4)$$

A generalized CA TSD construction, analogous to that in eq. 2, can be formulated. Letting $P(\psi|t_i, \mathbf{T})$ be the distribution of individual TSD's (ψ's) between thought t_i and the CA thinking $\mathbf{T}$, the generalized CA TSD is:

$$\psi(t_i, \mathbf{T}) = \int_0^1 f(\psi)P(\psi|t_i, \mathbf{T})d\psi \qquad (5)$$

Again, eq. 4 is recovered when $f(\psi) = \psi$. Note, this formulation in terms of an arbitrary convoluting function $f(\psi)$ enables a highly tunable speech function capable of selecting for thoughts with any arbitrarily chosen distribution of TSD's. Computing the speech function is then a matter of ranking thoughts by their TSD and speaking the highest ranked thought on behalf of the CA. With this, we define the generalized collective speech function as:

$$\Gamma(\mathbf{T}) \equiv \underset{t\in\mathbf{S}}{\operatorname{argmin}} \int_0^1 f(\psi)P(\psi|t_i, \mathbf{T})d\psi \qquad (6)$$

3.1 Data Collection

One might be first tempted to implement a Likert Scale where CA members are presented thoughts one by one and asked how well it agrees with what they are thinking. However, bias can arise which skew the reported results, such as those caused by cultural differences [10], thus a more robust method is adopted - pairwise comparison.

In a pairwise scheme each k^{th} CA member is presented with two thoughts, t_i and t_j, and asked which is closest to their thinking. Choosing t_i implies $\psi(t_i, \mathbf{T}^k) < \psi(t_j, \mathbf{T}^k)$. Data from each pairwise choice is stored in an array W_{ij} where $\mathbf{W} = 0$ at the start of the speaking cycle and W_{ij} is incremented if t_i is chosen over t_j. From this data set one can theoretically infer the TSD distribution functions using a standard MLE formulation. In practice, however, such a calculation is computationally expensive and does not (yet) lend itself to real-time computation. Thus, an easily computable TSD function is desirable. Assuming no correlations between the TSDs of two randomly sampled thoughts, letting $f(\psi) = \psi$, and noting that $\psi(t, \mathbf{T})$ can be remapped on the $[0, 1]$ interval with any function that preserves ordering; an approximate TSD, $\Psi(t_i|\mathbf{W})$, can be calculated for each thought directly from $\mathbf{W}$ via:

$$\Psi(t_i|\mathbf{W}) = \frac{1}{M}\sum_{j=1}^{M}\frac{W_{ij} + 1}{W_{ij} + W_{ji} + 2} \qquad (7)$$

Where M is the total number of thoughts entered by CA members. With this formulation $\Psi(t_i|\mathbf{W})$ is then an approximation of the probability that thought t_i would be chosen over any randomly selected thought [11].

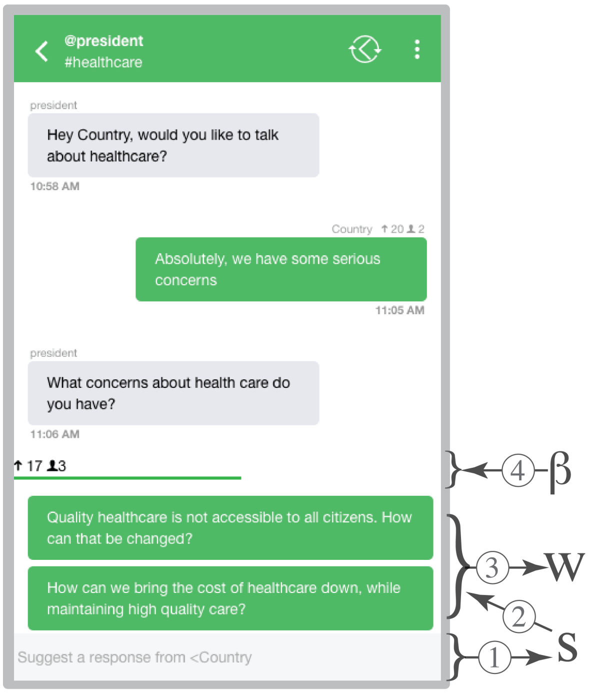
3.2 Conversation

Conversation is full duplex in nature - that is, participating agents must be able to respond dynamically to each other and react in real-time to thoughts being exchanged. For this reason, asynchronous CA speech is desired. To this ends, a speech cycle metric β must be considered which follows the progress of the CA speech cycle and crosses a threshold to mark its completion, at which time the top thought, t_p, is spoken. Some candidates for this are 1) the TSD of the top thought, $\Psi(t_p)$; 2) the uncertainty of the TSD of the top thought, $\frac{1}{M}\sum_{j=1}^{M}\frac{(W_{pj}+1)(W_{pj}+2)}{(W_{pj}+W_{jp}+2)(W_{pj}+W_{jp}+3)}$; 3) the average number of pair-wise comparisons per person 4) the ratio of pair-wise comparisons to the a^{th} power of the number of thoughts, $\frac{\sum_{ij}^{M} W_{ij}}{M^a}$; 5) the square of the average time derivative of all TSD's, $\frac{1}{M\Delta\tau}\sum_{i=1}^{M}(\Psi(t_i,\mathbf{W})|_\tau - \Psi(t_i,\mathbf{W})|_{\tau+\Delta\tau})^2$ and 6) the amount of time which has elapsed since the speech cycle began. Certainly, other metrics can be considered, and there is much room here for future work.

Bringing collective speech to the masses requires an internet based conversation platform which implements the model described above. With this in mind we have developed remesh, a communications platform that facilitates full-duplex text-based conversations over the internet between three basic configurations of agents: an individual agent (person) with another individual agent –classical text-based conversation–, an individual agent with a collective agent, and a collective agent with another collective agent. Built to appear and function as a chat app (shown in Figure 1), remesh implements the aforementioned model and is capable of supporting various speech cycle metrics and thought ranking schemes based on eq. 6.

4 Conclusion

This letter has outlined a general model for collective speech based on a simple experiential model for individual speech. While this model serves as a first step, there are many questions to be explored in future work. What constitutes an optimal TSD convoluting function $f(\psi)$? What speech cycle metric best tracks the speech cycle, and produces optimal collective speech? How can optimal collective speech be quantified? How can thoughts be sampled to achieve the best statistical sampling with the fewest comparisons? How can the members of a collective agent be chosen to produce super-intelligent speech on specific topics? How can CA members be incentivized to think on behalf of the group rather than the self? Are the myriad conversation based psychological tests [12] used to evaluate individuals valid for a CA? How can collective speech improve communication and decision making in corporate organizations? What role can collective speech play in governance, or inter-group conflict resolution? How can collective speech impact the organizational underpinnings of society at large? The remesh platform was built to serve as a springboard for exploring these questions and more, as well as to empower existing groups to speak for themselves.



Hey Country, would you like to talk about healthcare?

Absolutely, we have some serious concerns

What concerns about health care do you have?

Quality healthcare is not accessible to all citizens. How can that be changed?

How can we bring the cost of healthcare down, while maintaining high quality care?

Suggest a response from <Country

Figure 1: Mobile user interface for participating in collective speech within the remesh chat application. 1) Suggested messages are entered into a text box on the bottom of the screen which are added to the set of speakable thoughts **S**. 2) Two speakable thoughts are pulled from **S** and displayed to the participant. 3) The participant chooses which of the two thoughts presented is closest to their thinking and the choice is registered in **W**. 4) A progress bar displays the current value of the speech cycle metric β. When the progress bar fills (β crosses speaking threshold) the top thought, calculated from eq. 6, is sent on behalf of the CA and displayed within the conversation.

References

[1] Shonkoff, Jack P. (Editor); Phillips, Deborah A. (Editor); Committee on Integrating the Science of Early Childhood Development. *From neurons to neighborhoods: The science of early childhood development.* Washington, D.C.: National Academies Press; 2000.

[2] Turing, A.M. *Computing machinery and intelligence.* Mind, 59, 433-460. 1950.

[3] Lasecki, Walter S. *Powering Interactive Intelligent Systems with the Crowd* Proceedings of the Adjunct Publication of the 27th Annual ACM Symposium on User Interface Software and Technology, 21-24. 2014.

[4] Nowak, Stefanie, and Stefan Rüger.*How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation.* Proceedings of the international conference on Multimedia information retrieval. ACM, 2010.

[5] Lasecki W., Wesley R., Nichols J., Kulkarni A., Allen J., and Bigham J. *Chorus: A crowd-powered conversational assistant.* UIST. 2013.

[6] Meteyard L., Cuadrado S.R., Bahrami B., Vigliocco g., *Coming of age: A review of embodiment and the neuroscience of semantics*, Cortex, Volume 48, Issue 7, July–August 2012, Pages 788-804, ISSN 0010-9452

[7] Piccinini G., Scarantino A., *Information processing, computation, and cognition.* Journal Biological Physics (2011) 37:1–38

[8] Edelman, S. *On the nature of minds, or: truth and consequences.* Journal of Experimental and Theoretical Artificial Intelligence Vol. 20, No. 3, September 2008, 181–196

[9] Foo, N., et al. *Semantic distance in conceptual graphs.* Conceptual Structures: Current Research and Practice (1992): 149-154.

[10] Lee W., Jones S.,Mineyama Y., Zhang E. X., *Cultural Differences in Responses to a Likert Scale.*Research in Nursing and Health, 2002, 25, 295–306

[11] Salganik, Matthew J., and Karen EC Levy. *Wiki surveys: Open and quantifiable social data collection.* arXiv preprint arXiv:1202.0500 (2012).

[12] Groth-Marnat, Gary. *Handbook of Psychological Assessment.* Hoboken, NJ: John Wiley and Sons, 2003.