UTEP Technical Report UTEP-CS-09-19

Looking for Entropy Rate Constancy in Spoken Dialog

Alejandro Vega, Nigel G. Ward

Department of Computer Science
University of Texas at El Paso
500 West University Avenue
El Paso, TX 79968-0518

email: avega5@miners.utep.edu, nigel@utep.edu

June 18, 2009

The entropy constancy principle describes the tendency for information in language to be conveyed at a constant rate. We explore the possible role of this principle in spoken dialog, using the "summed entropy rate," that is, the sum of the entropies of the words of both speakers per second of time. Using the Switchboard corpus of casual dialogs and a standard ngram language model to estimate entropy, we examine patterns in entropy rate over time and the distribution of entropy across the two speakers. The results show effects that can be taken as support for the principle of constant entropy, but also indicate a need for better language models and better techniques for estimating non-lexical entropy.

Index Terms: distant context, overlapped speech, time into dialog, time into utterance

1 Background

The principle of constant entropy rate predicts that the entropy of communication will be constant over time, this being the most efficient way to use a communication channel [1]. There is evidence that this principle plays a role in human communication: in texts, in phonetic realization [2], in choice of syntactic forms [3], and in speaking rate across languages [4]. Other phenomena can also be explained using this principle: for example, the fact that "words subsequent to hesitations are less predictable than words uttered in fluent context" [5] can be seen as a strategy for balancing out high-entropy words with low-entropy silences and fillers, preserving the average entropy rate. Entropy constancy also relates to processing effort constancy: people seem to apply roughly the same effort level across time, making performance slower for more challenging tasks and texts [6, 7].

Previous work has considered written texts and monolog speech, but similar efficiency considerations apply to spoken dialog. Indeed, one could argue that dialogs should show greater

consistency of entropy than written text, as readers, unlike listeners, are free to change the speed at which they read, to adapt to changes in entropy rate. However one might also expect that that spoken dialog, as a spontaneous product of human minds, distractible and complex as they are, may exhibt less entropy constancy than planned text. In any case, there may be some role for entropy constancy in dialog, even if only as an ideal to which speakers strive. This paper is a first look at this possibility.

2 Defining and Estimating Entropy in Dialog

For two-party dialog it is necessary to consider the contributions of both speakers. We choose to focus on the entropy of the dialog as a whole, the "summed entropy," computed as the sum of the entropy in the productions of both speakers. Summed entropy approximates the total information content of a dialog. From the point of view of a participant, this is the total amount of information that he has to process, both as speaker (for his own utterances) and as listener (for the other's utterances).

We thus explore entropy constancy in dialog by considering the proposition that *in spoken dialog, there is a tendency for the rate of information conveyed, summed over the contributions of both participants, to be constant over time*; we call this the "summed entropy constancy principle". We choose to measure entropy rate relative to time, rather than to, say, words, because the same word can be spoken with different durations, and it seems that the speaking rate should affect the entropy rate.

Following Genzel and Charniak [1], our method of analysis is the examination of statistical patterns in entropy over time. For this we do not need highly accurate estimates of entropy (modulo the caveats below), so we used a standard trigram-based language model, that of the SRILM toolkit with default parameters. The training data consists of 3522 tracks from Switchboard [8], representing around 176 hours and 1.9 million words.

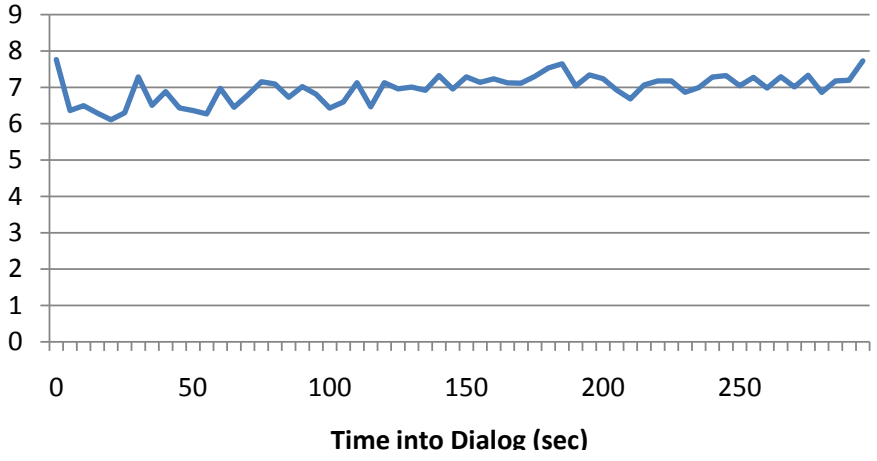

Figure 1: Entropy as a Function of Time Into Dialog, in bits per second

Handling out-of-vocabulary words is a messy issue, as always in entropy estimation. One option would be to estimate their entropy as the sum of the entropy of their letters [9], which seems overly harsh. At the other extreme, one could lump all oovs into one category. This

sometimes seems reasonable, for example in *"grew up in a small town, Canutillo"*, the word Canutillo is, for some purposes, equivalent to any unimportant place name not worth worrying about: it does not bear much information. We took a middle path: probabilities for words not seen in in the training data were crudely estimated as equal to those of words appearing only once in the training corpus, that is 1 / 1.9 million. Done without smoothing, this makes all our entropy estimates a trifle low.

Modeling the information content of silence is another problem — books have been written about the different meanings of silence. Following common practice, we simply used long silences (those over 1.2 seconds) to segment the corpus into separate utterances for the language model. Shorter silences were ignored. Thus, in either case, 'silence' tokens do not contribute to the entropy estimates.

Another problem is dealing with laughed words, as Switchboard has these labeled as distinct tokens, for example '[laughter-now].' We simply stripped off the 'laughter' annotations before computing the entropy.

Another problem comes from the fact that spoken words are not discrete events. Two tractable alternatives are to treat all the information content of a word as located at the moment of its onset, and to consider the information to be spread out evenly over the duration of the word. Gating experiments show that the truth lies between these extremes, but for simplicity we chose the first option, except for Table 1.

3 Entropy over the Course of a Dialog

As a preliminary step, we attempted to replicate Genzel and Charniak's finding [1] that in Treebank texts, the entropy of a sentence, computed without context, tends to increase with the sentence number. We expected to see a similar effect in spoken dialog, with the entropy estimates increasing with time into dialog. For each of 48 dialogs, for every 5-second timeslice, we computed the average entropy rate in that timeslice, by summing the entropies of all words starting in that slice and dividing by 5. Figure 1 shows the summed entropy rate in each slice, averaged across all dialogs.

There seems to be a tendency for a gradual increase in measured entropy over time. Genzel and Charniak's explanation for the increase they saw is also adequate to explain the effect in these dialogs. Since the estimate ignores such long-range information, being based only on local context, the entropy estimates will rise over time, due to the effects of the unmodeled context. For example, these dialogs typically start by establishing the topic (for example the prospects for the hometown team this season), and later utterances assume this, although this long-distance dependency is not modeled in our estimates. Thus, if the entropy is constant, the estimates are expected to increase over time, as is observed. Other factors may also be involved, as discussed below.

One unexpected behavior is the saliently high entropy in the first 5 seconds. This is probably because the words in this timespan are mostly greetings, which are not that common in the language as a whole, and hence are given high entropy estimates; even though they are, in reality, quite predictable.

	percentage of time	bits per second
silence	17%	–
speaker A only	36%	8.0
speaker B only	41%	8.1
A and B both speaking	6%	12.9
overall	100%	6.9

Table 1: Summed Entropy Rate vs. number of speakers

4 Entropy of Both Speakers

We then turned to test the principle more directly, by examining times of overlap. The principle predicts that the summed entropy at times of overlap will be the same as the entropy when only one person is speaking. At first glance this may seem implausible: if two people happen to talk at the same time, the summed entropy rate could logically be double that when only one person is talking. However there are also cognitive processing constraints in such situations: in general, people are able to either speak or listen, but not do both at the same time [10] (and the exceptions, notably back-channels, seem to typically be produced without much cognitive effort, to be perceived also almost effortlessly, and to convey relatively little information [11]).

To examine this prediction, we divided each dialog into regions based on who was speaking, and within each region summed the entropies of all words present, prorating the contributions of words falling partly within the region. (Prorating seemed more suitable here than using onset-based counting, but in fact both methods give essentially the same picture). The summed entropy rate was computed by dividing this sum by the length.

As seen in Table 1, the prediction was not confirmed: The entropy during times when both speakers were active was more than that during times with a single speaker. However some tendency towards entropy constancy is seen: the summed entropy was only 58% higher than for a single speaker, far less than double.

We wondered whether the picture could be confused by the inclusion of back-channels, which we expected to have low information content. We therefore computed the entropy rate separately for dual-speech regions lasting less than 500 milliseconds (mostly back-channels) and those more than 500 milliseconds (mostly not back-channels). In the shorter overlaps the summed entropy was 13.1, higher (surprisingly) than in the longer overlaps, 11.4. Thus the less-than-double entropy in overlapped speech is not attributable purely to back-channels.

A more detailed view, showing how two speakers in one dialog contribute to the summed entropy is shown in Figure 2. Here each dot represents the entropy contributions by the two speakers during a 1-second timeslice that included talk by both speakers. Overall the correlation is –0.258. Although the correlation is weak, the fact that it is negative is as predicted: when one speaker is conveying a lot of information, the other tends to be conveying little.

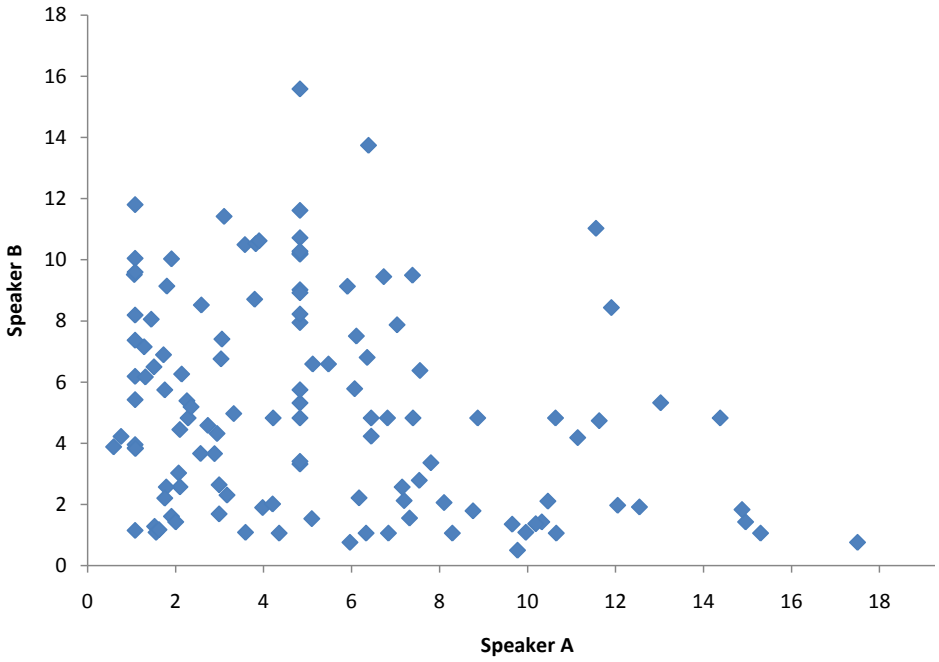

Figure 2: Relation between the Entropies of the Speakers during Simultaneous Talk

5 Entropy over an Utterance

In the interest of finding deviations from entropy constancy, we examined the role of time into utterance. We expected entropy to be lower around utterance starts, since at those times the interlocutor is likely to be *umm*-ing and *ah*-ing while formulating his utterance, and since gracefully taking the turn is also likely to be taking time and diverting cognitive effort. To test this we divided each utterance into 1-second slices. Figure 3 shows the result, averaged across all utterances in the 48 dialogs. The entropy is clearly not constant, although, contrary to prediction, the entropy in the first second was higher than elsewhere.

6 Discussion

Overall, graphs 1 and 3 are fairly flat, suggesting that the principle of constant entropy may hold in dialog to some extent. However we also observed systematic, significant, and substantial deviations, and in this section we consider some possible explanations for these before returning to discussion of the principle itself.

Looking again at Figure 3, the initial high entropy may be due merely to the weakness of the language model, which of course predicts the next word of the speaker without reference to information in the utterances of the other speaker, and thus unavoidably does a poor job at the utterance starts. More generally, the weakness is in using the summed entropy, rather than the joint entropy, as a metric; this is a limitation of the state of the art.

Looking again at Figure 1, the low entropy rate early in the dialogs could also be explained

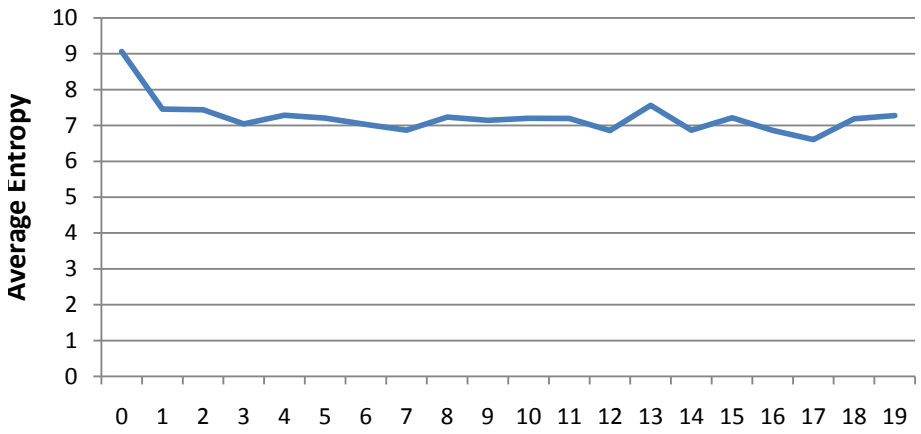

Figure 3: Entropy as a Function of Time Into Utterance

by the fact that, at the beginnings of these dialogs, the participants are also getting used to each other. The information conveyed during these activities may be largely non-semantic (involving gender, age, dialect, personality, dominance, etc.) and non-lexical (conveyed by voice features, pronunciation features, and prosody [12]). Quantifying and modeling of the entropy of such features would be useful for many reasons, and we are currently devising a way to do so by adapting Shannon's guessing game.

Thus there are many complications; so many that our method, of examining tendencies across many dialogs, is inadequate as a source of good evidence for or against entropy constancy in spoken dialog. If we choose to accept it anyway, as an *a priori* principle, the same complications severely limit its explanatory power for the kind of tendencies seen here.

Nevertheless, the entropy constancy hypothesis is compatible with our observations. Entropy constancy may have practical value, both for applying speech recognition to recorded dialogs and as a guiding principle for dialog management in spoken dialog systems. Future work could lay the groundwork for this by by studying entropy patterns in specific situations and contexts, and by examining the cognitive processes and communicative mechanisms that enable speakers to approach (and sometimes prevent them from approaching) the entropy constancy ideal.

Acknowledgments

We thank David G. Novick for insightful comments. This work was supported in part by the NSF under Grant No. 0415150 and by RDECOM via USC ICT.

References

[1] Dmitriy Genzel and Eugene Charniak. 2002. Entropy Rate Constancy in Text. *40th ACL*, 199-206.

[2] R. J. J. H. van Son and Louis C. W. Pols. 2003. How efficient is speech? Institute of Phonetic Sciences, University of Amsterdam, Proceedings 25, 171-184.

[3] Roger Levy and T. Florian Jaeger. 2006. Speakers Optimize Information Density through Syntactic Reduction. 20th Neural Information Processing Systems Conference.

[4] Francois Pellegrino, Christophe Coupe, and Egidio Marsico. 2007. An Information Theory-Based Approach to the Balance of Complexity between Phonetics, Phonology and Morphosyntax. presented at the 81st Annual Meeting of the Linguistic Society of America, Anaheim, CA, USA, 4-7 January 2007.

[5] Percy H. Tannenbaum and Fredrick Williams and Carolyn S. Hillier. 1965. Word Predictability in the Environments of Hesitations. *J. Verbal Learning and Verbal Behavior* (4) 134–140.

[6] Peter D. Bricker. 1955. Information Measurement and Reaction Time: A Review. in *Information Theory in Psychology*. Free Press. H. Quastler, ed. 350-359.

[7] Frank Keller. 2004. The Entropy Rate Principle as a Predictor of Processing Effort: An evaluation against eye-tracking data. EMNLP: 317-324.

[8] ISIP, Mississippi State University. 2003. Manually Corrected Switchboard Word Alignments. retrieved from http://www.ece.msstate.edu/research/ isip/projects/switchboard/

[9] Peter F. Brown, Vincent J. Della Pietra, Robert L. Mercer, Stephen A. Della Pietra, and Jennifer C. Lai. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics*, 11: pp 31-40.

[10] Joseph Jaffe. 1978. Parliamentary Procedure and the Brain. in *Nonverbal Behavior and Communication*. Aron W. Siegman and Stanley Feldstein, eds., Erlbaum. 55–66.

[11] Nigel Ward. 2006. Non-Lexical Conversational Sounds in American English. *Pragmatics and Cognition* (14): 113-184.

[12] A. M. Yaglom and I. M. Yaglom. 1983. *Probability and Information*. D. Reidel Publishing (Kluwer).