

CLASSIFICATION OF MEETINGS AND THEIR PARTICIPANTS

Cornelis Hoede and Xin Wang*

University of Twente
Dept. of Applied Mathematics, Faculty of EEMCS
P.O. Box 217,
7500 AE Enschede
The Netherlands

Abstract

On the basis of a coding of utterances we investigate ways to classify participants of a meeting. On the basis of a coding of states of a meeting activities during meetings are classified.

Key words: classification, meeting, graph, code

AMS classification: 05C99, 05E99, 91C99

1. INTRODUCTION

During meetings participants make statements, ask questions, give answers and produce various other types of utterances.

A first problem is whether one can distinguish between participants and whether relationships between them can be discovered. At our disposal was a report on an encoding system of meetings [2]. See Appendix 1 for a description of the various tags used for utterances and Appendix 2 for an example of a discussion encoded with these tags. A second problem is whether one can distinguish between activities during a meeting.

Op den Akker and Tommassen [1] distinguish 4 types of *activities*: silence S, brainstorm B, discussion D and presentation P. During all four activities the meeting is in one of 9 *states*. These states last for some time and occur in different ways during the activities. This information should enable to determine which activity takes place during a meeting at a certain time.

* On leave from Dalian Maritime University and Dalian University of Technology, Dalian, P.R. China

2. ANALYSIS OF A MEETING PROTOCOL

We will illustrate our methods by analyzing the protocol given in Appendix 2.

There are five participants c_1, c_2, c_3, c_4, c_5 . Their utterances can be counted for all types of tags. We will only consider statements S and questions Q . The frequency vectors are

$$\mathbf{S} = (12, 4, 5, 34, 5) \text{ and } \mathbf{Q} = (7, 3, 0, 1, 4).$$

It is already clear that the participants play different roles in the meeting. One way to make a classification on the basis of these two vectors is to distinguish e.g. high, medium and low numbers. Hoede and Wang[3] used auxiliary jurors to do this for a set of numbers like given here.

The simple outcome here is that c_4 scores high on statements, c_1 scores medium and c_2, c_3 and c_5 score low. On questions c_1 scores high, c_2 and c_5 score medium and c_3 and c_4 score low. Giving H, M and L as values the two vectors become

$$\mathbf{S} = (M, L, L, H, L) \text{ and } \mathbf{Q} = (H, M, L, L, M).$$

We see that c_2 and c_5 can be said to belong to the same class, low on statements and medium on questions, c_3 takes a somewhat offside position, whereas c_1 and c_4 are the most active participants.

In this way we can classify on the basis of any type of utterance. Possibly utterances are aggregated, like we did with statements and questions.

There is, however, another way to look at the protocol. If an utterance by one participant is followed by an utterance by another participant, this can be interpreted as a causal phenomenon; one utterance triggers another. A question followed by a statement is the standard example. When we count the pairs of utterances of different participants we can represent this by a directed labeled graph. An arc from c_i to c_j indicates that an utterance of c_i was followed by an utterance of c_j . The label indicates how often this happened. The result is given in Figure 1.

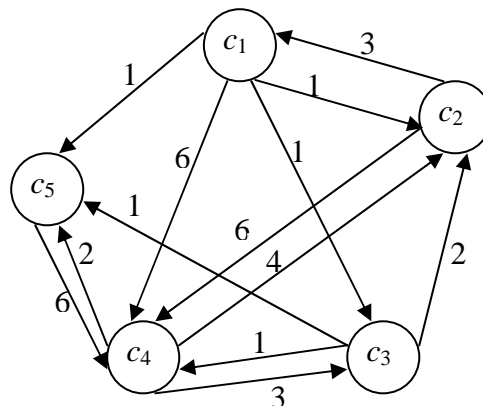


Figure 1
Reactions of participants on each other

The weighted indegrees are

$$\mathbf{id} = (3, 7, 4, 19, 4),$$

while the weighted outdegrees are

$$\mathbf{od} = (9, 9, 4, 9, 6).$$

c_4 shows high reaction, c_2 medium reaction, where as c_1, c_3 and c_5 show low reaction. c_1, c_2 and c_4 show high triggering, c_5 shows medium triggering and c_3 shows low triggering.

When we replace two opposite arcs by one arc in the direction of the arc with highest label and give that arc a label equal to the difference of the labels we obtain Figure 2.

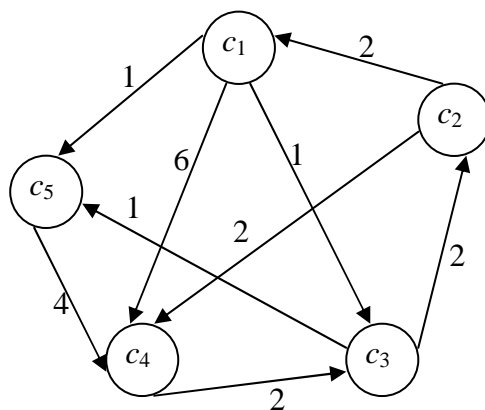


Figure 2

Reduced reactions of participants on each other

The indegree and outdegree vectors are now

$$\mathbf{id} = (2, 2, 3, 12, 2) \text{ and } \mathbf{od} = (8, 4, 3, 2, 4).$$

Like with the statements and questions, participants c_1 and c_4 come forward as most active participants, be it in completely different ways. c_4 by far is the most reactive participant whereas c_1 triggers most reactions. When we look a bit closer at the protocol we discover that c_4 is a woman called Carmen, see the first utterance of c_5 . c_1 clearly leads the discussion by his many explicit questions.

If we only focus on alternating utterances of pairs of participants, so e.g. c_5 - c_4 - c_5 - c_4 or c_4 - c_5 - c_4 - c_5 - c_4 , what hints at a real discussion between c_4 and c_5 started by c_5 respectively c_4 , we can give a “discussion graph”, as in Figure 3.

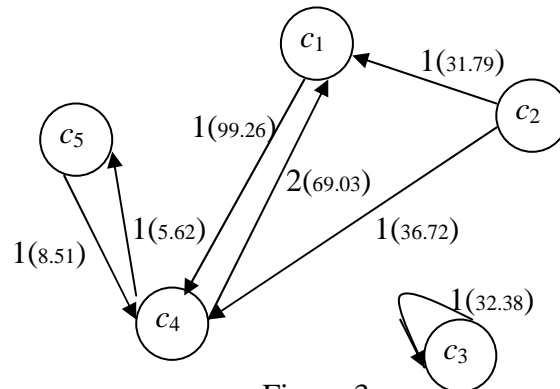


Figure 3
Discussion graph

The arcs are oriented from the participant that starts a discussion. The number besides the arc is the number of times conversation between the participants happened and the number in the brackets is the time the conversation lasted. The meeting started with a monolog of c_3 , that we indicated by a loop! The lady c_4 really comes forward as the central figure of the meeting from both the numbers of conversations that happened and the time they lasted. The time percentage of the conversation that c_4 was involved in is about 88%.

Another thing we want to point at is that although the number of the conversations related with somebody is quite large, the time percentage is possibly relatively small. Then we can not say that this person is the really central figure of the meeting. So we think that combining the duration of a speaker’s total meaningful talking (here we mean the talking except “yeah”, “well”, “so”, “oh”, e.g.) and the number of conversations he is involved in can help us to get the real central figure in a meeting.

3. ANALYSIS OF A MEETING WITH RESPECT TO ACTIVITIES

The conversation states in which a meeting can be, according to Op den Akker and Tommassen[1], are given in Table I, together with the percentages of the total time of 8 meetings, during which these states occurred. The description of the types of states can be found in [1].

Table I
Distribution of conversational states

State	Percentage
Silent	25
Only stalls	6.7
Only backchannels	0.58
Only stall and backchannel	0.15
Single speaker	54
Stall	4.9
Backchannel	1.5
Speaker overlap	6.7
Other	1.1

The four types of activities distinguished turned out to each have a distribution of the time over the nine states distinguished. Table II gives these distributions of conversational states over the activities as vectors of 9 elements.

Table II
Distribution of the conversational states among the activities

state	Distribution of the state
Silence	(63, 7.8, 0.65, 0.04, 23, 2.5, 0.2, 2.8, 0.48)
Brainstorming	(14, 7.5, 0.62, 0.36, 55, 7.8, 2.0, 11, 2.2)
Discussion	(13, 5.3, 0.47, 0.12, 61, 6.6, 2.0, 9.9, 1.3)
Presentation	(10, 6.1, 0.60, 0.08, 76, 2.6, 1.7, 2.7, 0.16)

These numbers may be seen as averages, obtained from annotating the 8 meetings. The problem we are facing is to determine the activity going on during a meeting on the basis of observation of the states and how long these states pertain.

Given the observations over a certain region of time, we obtain a 9-element vector from which we have to decide upon the activity going on. One of the obvious ways to classify a meeting period is to use a decision tree derived from the training examples given by the annotations of the 8 meetings. In [1] three other ways of classification are described as well, one of which is by using neural networks.

A decision tree will have to split according to 9 attributes, with at least 2 values per attribute. Distinguishing values H, M and L for each attribute, the full decision tree would have $3^9=19683$ end nodes. Distinguishing only 2 values, say H and L, still $2^9=512$ end nodes would occur in the full decision tree.

We want to propose an alternative way of classifying a meeting on activities going on. The basic idea is to see Table II as describing code words. Observing a certain period of a meeting gives a “message” vector that has a certain “distance” with respect to the four code words. The classification then simply takes place by determining which of the four code words is closest. We will show the usefulness of this idea in a reduced analysis of Table II.

First we replace the numbers by H, M, and L, applying the following method. The lowest and highest average for a conversational state determine an interval [a, b]. We choose $a+(1/4)(b-a)$ and $a+(3/4)(b-a)$ as boundaries for L, M and H values. The idea behind this is that, as we consider averages, L-values are found around a and H-values around b. Moreover the intervals for these values should be more or less the same in length. We obtain Table III.

Table III
Reference activities encoded

Activity	Values of the states
Silence	(H, H, H, L, L, L, L, L, L)
Brainstorming	(L, H, H, H, M, H, H, H, H)
Discussion	(L, L, M, L, M, H, H, H, H)
Presentation	(L, M, M, L, H, L, H, L, L)

A further simplification is to focus on the higher values in Table II. The five states we chose as relevant are: silent (1), only stalls(2), single speaker(3), stall (4) and overlap (5), moreover we replace H, M and L by 2, 1 and 0. This yields the following code words;

Table IV
Reference activities as code words

Activity	Code word
Silence(S)	(2, 2, 0, 0, 0)
Brainstorming(B)	(0, 2, 1, 2, 2)
Discussion(D)	(0, 0, 1, 2, 2)
Presentation(P)	(0, 1, 2, 0, 0)

We now have to derive a distance functional, first for our four code words.

We calculate the differences of corresponding vector elements and sum them, so S and D are at maximum distance 10, whereas B and D have only distance 3, indicating that there is not much difference between these two activities. For an *observation vector*, with percentages as elements, we have to translate the vector to the same format. This can be done by defining for each of the five states boundaries for the attribute values H, M and L. For example, the percentages for single speaker are 23, 55, 61 and 76. Within the interval [23, 76] we choose $23+53/4=36.25$ and $23+159/4=62.75$ as boundaries.

The observation gives a message vector and the classification is by the closest code word. This is quite easy. However, a problem comes forward when we ask which period is to be considered, when during a meeting the activities vary. The meeting may start with a period of silence, followed by a period of presentation, leading to a period of discussion, getting back to a period of presentation again, followed by a brainstorm.

The message vector should “move” from the neighbourhood of S to that of P, to that of D, to that of P again and finally to that of B. This brings in a “dynamic” aspect. Before showing how to handle this we first want to give an example.

Let a meeting be in the following sequence of states: $1 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 3 \rightarrow 2$. The time periods are assumed to be such that the partial periods ending with a change of state have the following message vectors:

- 1 : (2, 0, 0, 0,0) \Rightarrow S
- 1 \rightarrow 3 : (2, 0, 1, 0,0) \Rightarrow S
- 1 \rightarrow 3 \rightarrow 4: (1, 0, 0, 2, 0) \Rightarrow D
- 1 \rightarrow 3 \rightarrow 4 \rightarrow 5: (0, 0, 0, 2, 2) \Rightarrow D
- 1 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1: (0, 0, 1, 1, 1) \Rightarrow D
- 1 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 3: (0, 0, 1, 1, 1) \Rightarrow D
- 1 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 1 \rightarrow 3 \rightarrow 2: (0, 1, 1, 1, 1) \Rightarrow D, B or P.

The resulting classifications follow from calculation of distances, the message vector starts in the neighbourhood of S, goes to the neighbourhood of D, and at the end has equal minimum distances to the code words of D, B and P. An activity of silence seems followed by an activity of discussion. However, the observation concerns an ever growing time interval. In order to classify the current activity at a certain moment it is more natural to use that part of the sequence of states that ends at the considered moment. The problem that comes forward then is how far back the begin point of the partial sequence is to be chosen. A way to handle this problem is to consider a number of partial sequences from

2, 3 →2, 1→3→2, 5→1→3→2, 4→5→1→3→2, 3→4→5 →1→3→2 and 1→3→4→5→1→ 3→2.

Suppose these partial sequences are classified as S, P, P, B, B, P and B or P.

The picture now completely changes. At the end of state 2 most of the classifications give P. As the last three partial sequences give B, P and P we may conclude that **at the considered moment** the activity was a presentation. The partial sequences till the moment after state 5 might give classification B.

4. ANOTHER WAY TO DECIDE ON THE ACTIVITY

We propose another way to classify the ongoing activity by listing all the possibilities of the order of a small number of last states. For example, we take 3 states here, e.g.

1→2→3. Then

we consider the average total time intervals [25, 6.7, 54] from Table I: distribution of conversational states. Now we change these numbers into percentages [29.2, 7.8, 63], in order to compute the values in Table III: Reference activities encoded. Then we get 80 possibilities of records in Table V.

Table V
Activities decided by the order of the last 3 states

Records	States order	Activity	States order	Activity	States order	Activity	States order	Activity
1	123	P	124	S	134	D	245	B
2	132	P	125	S	135	D	254	B
3	213	P	142	S	143	D	324	B
4	231	P	152	S	145	D	325	B
5	234	P	214	S	153	D	342	B
6	235	P	215	S	154	D	352	B
7	243	P	241	S	314	D	425	B
8	253	P	251	S	315	D	452	B
9	312	P	412	S	341	D	524	B
10	321	P	421	S	345	D	542	B
11	423	P	512	S	351	D	242	B
12	432	P	521	S	354	D	252	B
13	523	P	121	S	413	D	424	B
14	532	P	141	S	415	D	525	B
15	131	P	151	S	451	D	545	B

16	232	P	212	S	513	D		
17	313	P	414	S	514	D		
18	323	P	515	S	531	D		
19	343	P			534	D		
20	353	P			541	D		
21	535	P			543	D		
22					431	D		
23					435	D		
24					453	D		
25					434	D		
26					454	D		

If we observe Table V, there are some interesting rules. All the sequences in activity P have 3 while those in S have no 3; similarly, all the sequences in S have 1 and those in B have no 1 at all; as for activity D, 2 never appears and it must have either 4 or 5. All these rules really agree with our intuitions.

We can also consider the order of 2 states or 1 state to decide on the activity. Then we get Table VI.

Table VI
Activities decided by the order of the last 2 or 1 states

Records	S.O	A	Records	S.O	A	Records	S.O	A	Records	S.O	A	Records	S.O	A
1	21	S	5	12	S	9	13	P	13	14	S	17	15	S
2	31	P	6	32	P	10	23	P	14	24	B	18	25	B
3	41	S	7	42	B	11	43	P	15	34	P	19	35	P
4	51	S	8	52	B	12	53	P	16	54	D	20	45	D
21	1	S	22	2	S	23	3	P	24	4	D	25	5	D

5. TESTING THE CLASSIFICATION

We now want to determine the correctness rate of this way of classifying activities. Before testing this way of classifying activities on real life data in a future paper we construct some artificial data. Table V and VI were calculated on the basis of Table I, that gave overall averages of time used for the 9 states a meeting could be in. Suppose one wants to find out which activity is taking place by listening in for some time, say at most five minutes, at some meeting. This “measuring” starts at some moment, during the time period the meeting is in some state. If in the next five minutes the state does not change, we can use Table VI. Five minutes of silence would lead to the conclusion that the activity is silence and five minutes of single speaker to the conclusion that a presentation is going on.

In order to point out a major difficulty of our classification method, let us consider the states order 313: a single speaker state followed by silence and a single speaker again. As soon as the change from 1 to 3 is perceived, we could use Table V and classify as P, assuming that for both single speaker observations the average time can be expected.

However, suppose that the observed period was a distribution of percentages of time of 5%, 90%, 5% for 3, 1 and 3 respectively, then it seems more natural to conclude that the activity is S. In the observed period state 1 must be encoded as H and state 3 as L, and the period has encoding (H, L, L, L, L) or (2, 0, 0, 0, 0). From Table IV we see that S has indeed the smallest distance to the observed states order, when these percentages are taken into account. When the states 3 last longer, say 50%, 45% and 5% are measured in the period, then 3 is encoded as M and 1 as H. The code word for the period becomes (2, 0, 1, 0, 0), having distance 3 to S and distance 4 to P, so still the activity is classified as silence. As states 3 are expected to last longer, see Table I, the way to measure seems to be as follows: starting in some state, the next three states are measured, also in duration. This will ask for four changes of state. If the order 313 is preceded by a state $S_1 \neq 3$ and followed by a state $S_2 \neq 3$, then in the order $S_1 \rightarrow 3 \rightarrow 1 \rightarrow 3 \rightarrow S_2$ the measurement starts during S_1 and ends during S_2 , thus making sure that the percentages can be determined and the encoding can take place. As said before, it may happen that within the prescribed five minutes less than four changes take place. In case only three changes take place, we may have a states order $S_1 \rightarrow 3 \rightarrow 1 \rightarrow S_2$, conclude that $3 \rightarrow 1$ is measured, and encode such a period. In case only two changes take place, we may have states order $S_1 \rightarrow 1 \rightarrow S_2$, leading to the conclusion that the activity going on is S. The same holds in case we perceive $S_1 \rightarrow 1$ during the five minutes or just 1, in case there are no changes.

We now want to see whether this more sophisticated way of measuring leads to an improvement of the correctness rate of the classification. For this we construct an artificial meeting in the following way. Let the activity in the meeting be a discussion. From Table II we see that the distribution over the five states is (13, 5.3, 61, 6.6, 9.9).

These numbers roughly are 2×6 , 1×6 , 10×6 , 1×6 , 2×6 , with proportional relations 2:1:10:1:2. We now consider 16 time intervals, 2 in state 1, 1 in states 2, 10 in states 3, 1 in state 4 and 2 in state 5. Any meeting constructed from these 16 intervals has to be classified as D, the time intervals states may occur in any order e.g.

D: 1331323355433333.

There are $16!/(2!1!10!1!2!)=1.441440$ different orderings.

We now have an example of a discussion and simulate measurements.

First we listen in on D during some time interval and stop when two changes have taken place. We suppose any time interval could be the one in which the measurement starts. We pose no time limit. The states orders and classifications via Tables V and VI found are

1331	P	S_1 3313	S_2	(0,0,2,0,0)	P
3313	P	S_1 132	S_2	(1,2,0,0,0)	S
313	P	S_1 132	S_2	(1,2,0,0,0)	S
132	P	S_1 3233	S_2	(0,2,2,0,0)	P
323	P	S_1 23355	S_2	(0,2,0,0,2)	B
2335	P	S_1 33554	S_2	(0,0,0,2,2)	D
33554	D	S_1 55433333		(0,0,1,2,2)	B

3554	D	S ₁ 55433333	(0,0,1,2,2) B
5543	D	S ₁ 433333	(0,0,2,2,0) D
543	D	S ₁ 433333	(0,0,2,2,0) D
433333	P	S ₁ 333333	(0,0,2,0,0) P
33333	P	S ₁ 3333	(0,0,2,0,0) P
3333	P	S ₁ 333	(0,0,2,0,0) P
333	P	S ₁ 33	(0,0,2,0,0) P
33	P	S ₁ 3	(0,0,2,0,0) P
3	P		

The measurements in most cases classify the ongoing activity as presentation. Only if the measurement takes place in the middle, what is going on is classified as discussion. The more complicated classification gives more differentiation. Whereas the first method describes 3 periods of activities, P, D and P, the second method indicates a period of silence and presentation, followed by a period of discussion and brain storming, ending with a period of presentation.

Applying the same procedure to a meeting that overall is of type silence we have from Table II the distribution:

$$(63, 7.8, 23, 2.5, 2.8)$$

Roughly these numbers show proportionality 24:3:9:1:1. Hence we consider 38 time intervals, 24 in state 1, 3 in state 2, 9 in state 3 and 1 in states 4 and 5. A random permutation may look like:

S: 11111111311312133411231511231311311113

The first method now gives relatively many times a classification as presentation. The second method, with S in one of the eight first time intervals, gives S₁3113S₂, so measures 3113 with encoding (2, 0, 1, 0, 0) with distance 3 to S and distance 4 to B, see Table IV, so the classification gives S. Measuring starting in the last interval in state 2 gives e.g. S₁ 313 S₂, encoding (1, 0, 1, 0, 0) and classification P, but from there on we find:

$$\begin{array}{llll}
S_1 1311 S_2 \Rightarrow S & S_1 11113 \Rightarrow S & S_1 3113 S_2 \Rightarrow S & S_1 1113 \Rightarrow S \\
S_1 1131111 \Rightarrow S & S_1 113 \Rightarrow S & S_1 311113 \Rightarrow S & S_1 13 \Rightarrow S \\
S_1 311113 \Rightarrow S & S_1 3 \Rightarrow P. & &
\end{array}$$

So silence as classification in almost all cases.

6. DISCUSSION

From the results presented in Section 5 we conclude the following.

1. Using the more sophisticated measurements, so measuring actual duration of states, is to be recommended for classifying ongoing activities. The discussion example showed that much more differentiation is made than by using Tables V and VI, that are based on assumed average durations. The silence example showed that the differences in average time tend to give more classifications as ongoing presentations, when using the first method.
2. Measuring the activity for a short time may lead to a classification that deviates from the overall classification. In the discussion example we considered measurements starting during each of the 16 constructed time intervals. Although the overall activity is a discussion, this was only measured in 4 out of 16 cases by the first method and in only 3 out of 15 cases by the second method.
3. It only makes sense to make a statement about a certain time interval. The resulting classification refers to that time interval and can not be used to infer a classification of the overall meeting.
In comparing our classification method we should use classifications of meetings as activities by human classifiers, and calculate our classification for those overall meetings.

REFERENCES

- [1] R. op den Akker and P. Tommassen, *Classification of Meeting Activities based on Conversation State Sequences and Speaker Activities*, Department of Computer Science, University of Twente, The Netherlands, Preprint(2006).
- [2] R. Dhillon, S. Bhagat, H. Carvey and E. Shriberg, *Meeting Recorder Project: Dialog Act Labeling Guide*, Department of Computer Science, University of Twente, The Netherlands, (2003).
- [3] Cornelis Hoede and Xin Wang, *On Fuzzy Concepts*, Memorandum No. 1814, Department of Applied Mathematics, University of Twente, The Netherlands, (2006).

Appendix 1: Meeting Recorder DA (MRDA) Tagset

Group 1: Statements

s Statement

Group 2: Questions

qy Y/N Question

qw Wh-Question

qr Or Question

qrr Or Clause After Y/N Question

qo Open-ended Question

qh Rhetorical Question

Group 3: Floor Mechanisms

fg Floor Grabber

fh Floor Holder

h Hold

Group 4: Backchannels and Acknowledgements

b Backchannel

bk Acknowledgement

ba Assessment/Appreciation

bh Rhetorical Question Backchannel

Group 5: Responses

Positive

aa Accept

aap Partial Accept

na Affirmative Answer

Negative

ar Reject

arp Partial Reject

nd Dispreferred Answer

ng Negative Answer

Uncertain

am Maybe

no No Knowledge

Group 6: Action Motivators

co Command

cs Suggestion

cc Commitment

Group 7: Checks

f "Follow Me"

br Repetition Request

bu Understanding Check

Group 8: Restated Information

Repetition

r Repeat

m Mimic

bs Summary

Correction

bc Correct Misspeaking

bsc Self-Correct Misspeaking

Group 9: Supportive Functions

df Defending/Explanation

e Elaboration

2 Collaborative Completion

Group 10: Politeness Mechanisms

bd Downplayer

by Sympathy

fa Apology

ft Thanks

fw Welcome

Group 11: Further Descriptions

fe Exclamation

t About-Task

tc Topic Change

j Joke

t1 Self Talk

t3 Third Party Talk

d Declarative Question

g Tag Question

rt Rising Tone

Group 12: Disruption Forms

% Indecipherable

%- Interrupted

%-- Abandoned

x Nonspeech

Group 13: Nonlabeled

z Nonlabeled

APPENDIX 2: LABELED MEETING SAMPLE

A labeled five-minute portion of Bro021 is shown below. Included are start and endtimes, channel numbers, DAs, adjacency pairs, and the corresponding portions of the transcript.

1828.250-1832.820	c3	s		i like plugged some groupings for computing this eigen- - uh uh
1832.820-1839.250	c3	s		uh s- - values and eigenvectors . so just - i just some small block of things which i needed to put together for the subspace approach .
1839.250-1845.680	c3	s		and i'm in the process of like building up that stuff .
1846.670-1849.080	c3	fh		and um ==
1850.400-1852.790	c3	fh		uh - yeah .
1854.120-1856.580	c3	s		i guess - yep i guess that's it .
1856.580-1859.040	c3	s		and uh th- - th- - that's where i am right now .
1859.620-1860.630	c3	fh		so .
1861.560-1863.000	c5	qo^tc	1a	oh how about you carmen ?
1862.830-1865.740	c4	s	1b	huh i'm working with v t s .
1866.330-1869.160	c4	fh s		um i do several experiment with the spanish database first .
1869.150-1873.400	c4	s^e	2a	only with v t s and nothing more .
276.050-279.990	c2	s^e		to adapt more quickly to the r- - something that's closer to the right mean .
1875.520-1876.580	c4	s^e		no l d a .
1873.400-1875.520	c4	s^e		not v a d .
1876.580-1877.640	c4	s^e		nothing more .
1877.030-1878.270	c5	qw^rt	2b.3a	what - what is v t s again ?
1878.070-1881.140	c4	s	3b.4a	uh vectorial taylor series .
1878.320-1879.090	c3	%-		new ==
1880.420-1881.070	c5	s^bk	4b	oh yes .
1881.070-1881.710	c5	s^aa	4b+	right right .
1881.350-1883.060	c4	s		to remove the noise too .
1882.530-1885.350	c5	s	5a	i think i ask you that every single meeting .
1885.350-1886.750	c5	qy^g	5a+	don't i ?

1884.860-1885.590	c4	qw^br	5b.6a	what ?
1886.750-1888.160	c5	s	6b.7a	i ask you that question every meeting .
1887.310-1888.120	c4	s^aa	7b-1	yeah .
1888.120-1888.930	c4	%-		if - well ==
1888.080-1890.790	c1	s^j	7b-2.8a	so that'd be good from - for analysis .
1890.790-1892.140	c1	s^df^j	7b-2+.8a+	it's good to have some uh cases of the same utterance at different - different times .
1892.140-1893.490	c1	fh		yeah .
1891.680-1893.200	c5	s^bk	8b	yeah .
1893.200-1894.720	c5	qw^j	8b+.9a	what is v t s ?
1895.100-1896.260	c4	s^m	9b	v t s .
1896.260-1897.410	c4	s.%--		i'm sor- ==
1897.410-1898.980	c4	s.%--		well um the question is that ==
1898.980-1900.540	c4	fh		well .
1900.540-1903.300	c4	s		remove some noise but not too much .
1903.700-1909.290	c4	fh s		and when we put the m- - m- - the them - v a d the result is better .
1909.290-1915.030	c4	s		and we put everything the result is better .
1915.030-1920.770	c4	s	10a	but it's not better than the result that we have without v t s .
1921.110-1921.780	c4	s^ar		no no .
1923.210-1924.060	c1	s^bk	10b	i see .
1924.060-1930.290	c1	s.%--	11a	so that given that you're using the v a d also the effect of the v t s is not so far ==
1929.630-1930.270	c4	s^na	11b	is not .
1930.780-1934.640	c1	qw^rt	12a	do you - how much of that do you think is due to just the particular implementation and how much you're adjusting it ?
1934.640-1938.490	c1	qw.%--	12a+	or how much do you think is intrinsic to ?==
1936.770-1937.830	c4	s^no	12b	pfft i don't know .
1937.830-1938.880	c4	s^df.%--	12b+	because ==
1938.880-1940.500	c4	fh		hhh ==
1939.210-1941.350	c2	qy	13a	are you still using only the ten first frame for noise estimation ?
1941.350-1943.490	c2	qrr.%--		or ?==
1944.260-1953.610	c4	h s^rt	13b	uh i do the experiment using

				only the f- - onl- - uh to use on- - only one fair estimation of the noise .
1944.890-1946.040	c2	qrr.%--		or i- ?==
1948.290-1948.820	c2	b		yeah .
1949.670-1950.580	c2	b		huh .
1953.610-1961.850	c4	s	13b+	and also i did some experiment uh doing um a lying estimation of the noise .
1962.430-1965.860	c4	s.%--		and well it's a little bit better but not ==
1966.550-1967.100	c4	x		n- ==
1967.920-1969.610	c2	s^cs		maybe you have to standardize this thing also .
1970.450-1974.600	c2	s^df.%--		because all the thing that you are testing use a different ==
1969.610-1970.450	c2	s^e		noise estimation .
1975.430-1975.930	c4	b		huh .
1975.490-1976.000	c3	b		huh .
1975.780-1978.860	c2	s^df		they all need some - some noise - noise spectra .
1978.860-1981.940	c2	s^df		but they use - every - all use a different one .
1976.720-1979.030	c4	s^ar s		no i do that two - t- - did two time .
1982.310-1983.860	c1	s		i have an idea .
1983.860-1985.620	c1	s.%--		if - if uh uh ==
1985.620-1986.500	c1	s^aa		y- - you're right .
1986.500-1987.380	c1	s		i mean each of these require this .
1987.380-2000.980	c1	qw^cs		um given that we're going to have for this test at least of - uh boundaries what if initially we start off by using known sections of nonspeech for the estimation ?
1999.540-2000.350	c4	b		uhhuh .
1999.630-2000.020	c2	b		uhhuh .
2003.140-2003.740	c1	qy^d^g^rt		right ?
2003.740-2005.860	c1	fh		s- - so e- - um ==
2003.760-2004.160	c2	b		yeah .
2004.160-2004.570	c2	b		uhhuh .
2005.860-2010.710	c1	s^df		first place i mean even if ultimately we wouldn't be given the boundaries uh this would be a good initial experiment to

2010.710-2015.930	c1	qw		separate out the effects of things . i mean how much is the poor you know relatively uh unhelpful result that you're getting in this or this or this ?
2015.930-2021.370	c1	qy		is due to some inherent limitation to the method for these tasks ?
2021.370-2031.420	c1	qw		and how much of it is just due to the fact that you're not accurately finding enough regions that - that are really n- - noise ?
2028.600-2029.070	c3	b		huh .
2030.230-2030.880	c4	b		uhhuh .
2030.780-2031.490	c2	b		uhhuh .
2032.080-2033.070	c1	fh		um ==
2033.070-2037.980	c1	s^df	14a	so maybe if you tested it using that you'd have more reliable stretches of nonspeech to do the estimation from .
2037.980-2042.900	c1	s	14a+	and see if that helps .
2042.880-2045.120	c4	s^bk	14b	yeah .
2045.120-2046.250	c4	s^tc		another thing is the them - the codebook .
2046.250-2047.370	c4	s^bsc		the initial codebook .
2047.370-2049.380	c4	s.%--		that maybe ==
2049.380-2050.380	c4	s		well it's too clean .
2050.380-2051.380	c4	fh		and ==
2051.240-2051.980	c1	b		uhhuh .
2051.380-2052.560	c4	s^df.%--		because it's a ==
2052.560-2053.150	c4	fh		i don't know .
2053.150-2053.740	c4	s.%--		the methods ==
2054.740-2058.370	c4	s^cs	15a	if you want you c- - i can say something about the method .
2058.420-2059.090	c1	s^aa	15b	uhhuh .
2059.380-2060.780	c4	s.%--		yeah in the ==
2065.040-2070.080	c4	s^df		because it's a little bit different of the other method .
2071.310-2072.790	c4	s.%--		well we have ==
2073.710-2088.990	c4	s		if this - if this is the noise signal uh in the log domain we have something like this .
2102.010-2103.390	c4	s		now we have something like this .
2103.390-2107.640	c4	s.%--		and the idea of these methods is

2107.640-2111.900	c4	qw	to n- - given a um == how do you say ?
2108.620-2110.040	c1	b	huh huh .
2111.900-2115.240	c4	s	i will read because it's better for my english .
2116.130-2117.780	c4	%--	i- - i- - given ==
2117.780-2120.610	c4	s	is the estimate of the p d f of the noise signal .
2120.610-2131.340	c4	s	when we have a - um a statistic of the clean speech and an statistic of the noisy speech .