Meeting Recorder Project: Hot Spot Labeling Guide

B. Wrede, S. Bhagat, R. Dhillon, E. Shriberg

TABLE OF CONTENTS

1. Introduction

The annotation of hot spots as part of this corpus was conducted as a modest added effort to dialog act annotations, because annotators had a few months left at the end of the overall project. We wanted some rather quick and general way to assess affect in the meetings, but knew that there is little evidence of the canonical emotions traditionally studied in the emotion literature. Even though ICSI meetings are natural, they do not have frequent instances of joy, sadness, fear, etc. Although there are many instances of propositional disagreement, which are marked in the dialog act annotations, it is rare to find someone openly displaying signs of anger or frustration in this data set. There are, however, instances of increased affective and/or interactive involvement, particularly for disagreements and humor. We call these instances hot spots, and have labeled them for the full 75 meetings.

An important property of the annotation scheme for computational purposes is that hot spots are annotated on top of dialog acts, to enable easy alignment of the two types of information. That is, dialog acts are the smallest units on which hot spots are labeled (in the rare cases in which hot spot quality changes within a dialog act, special labels are used to allow recovery of the information.) Hot spots have both an extent and an internal structure, as well as a level and a type. This allows the researcher to map from the structured annotations to whatever type of annotation is of most interest to their research. For example, one can look only at beginnings or ends, or at only certain types or levels. One can also use only the 'peaks' in the hotspots, or can propagate peak annotations to all dialog acts within the begin-end region of that hot spot, and so on, depending on the interest at hand.

2. Definition of Hot Spots

Hot spots are parts in conversations that stand out from the rest of the conversation in that

- the participants are more involved (emotionally or "interactively")

- there is a higher degree of interaction by participants who are trying to get the floor

3. Format of a Hot Spot

A hot spot is comprised of a starter, any sequential utterances, and the finisher of the hot spot.

A **complete hot spot label** has the following components:

> <Number>=<Level of Hotness><Degree>=<Type[(s);secondary type(s)]>.<Utterance Tag>

The above complete hot spot label is marked at the starter of a hot spot. The labels following the start of a hot spot up until the finish of the hot spot receive only an utterance tag if necessary. The hot spot label marked only at the start applies to all the utterances belonging to the hot spot. However, there are exceptions where utterances following the start do receive a <Number> tag. These cases are explained in the following section.

4. Number

Hot spots are given a numerical value in order to differentiate between hot spots. They are numbered in ascending order in increments of 10. This system of numbering hot spots in multiples of 10 allows for intermediate hot spots that have been overlooked to later receive whole number values. For instance, if between hot spots 10 and 20 another hot spot is found, that hot spot can receive a numerical value of 15.

Because of how the time bins are ordered (by end time), there may be hot spots in which certain utterances belonging to a hot spot temporally are actually listed before the complete hot spot label at the start of the hot spot. In such cases, these time-misaligned utterances should be labeled with the numerical value of the hot spot to which they belong to.

Example 1: Bmr003

1722.55,1726.16,Bmr003-c4,s,71a+,**120=l=a;c.a**,,if - if that could handle speaker identification that's a big deal .

1724.19,1724.99,Bmr003-c1,%-,,,,well it w- ==

1728.02,1728.67,Bmr003-c3,s^co,72b-1,,,well use it .

1728.1,1728.56,Bmr003-c2,s^bk,72b-2,,,right .

1728.29,1728.63,Bmr003-c4,b,,,,o_k .

1728.67,1730.91,Bmr003-c3,s^df,72b-1+,**b.e**,,yeah that's why we s- - bought the expensive microphones .

```
1729.21,1730.95,Bmr003-c4,s^bk|s^ba,72b-3,,,yeah | i mean that's a nice feature.

1730.49,1730.75,Bmr003-c0,b,,,,yep .

1730.8,1731.57,Bmr003-c1,b,,120b,,<i> <think> <it's> <pretty> <easy> <when>
<we've> <got> <separate> <channels> yeah yeah .
```

Do note also that a hot spot can never occur *within* a larger hot spot. Hot spots always should be labeled in a consecutive order.

5. Level of Hotness

The level of hotness has three classes: <l> for lukewarm, <w> for warm, and <h> for hot. A hot spot with a level of hotness of <l> indicates that the "hot" utterances in the hot spot are mildly involved. Lukewarm hot spots usually contain minimal laughter or involvement. Lukewarm hot spots tend to be difficult to detect. A hot spot with a level of hotness of <w> indicates that the "hot" utterances in the hot spot are markedly involved. While a hot spot with a level of hotness of <l> is difficult to detect, a hot spot with a level of hotness of <w> is easily perceptible. A hot spot with a level of hotness of <h> indicates that the "hot" utterances in the hot spot are heavily involved and have a high degree of intensity. To differentiate between the levels <w> and <h>, the primary distinction is that hot spots marked with <h> are far more uproarious and louder than those marked with <w>.

Additionally, hotness is speaker-dependent and only pertains to how an utterance is uttered as opposed to surrounding utterances or laughter. Speaker-dependent patterns are important when, for instance, considering the utterances of a generally enthusiastic speaker. For such a speaker's utterances to be indicative of a hot spot, the level of involvement or emotion on the part of the speaker must be higher than it normally is. Similarly, if a speaker has a tendency to speak in a calm manner, then there may very well be cases in which this speaker's hotness level is actually equivalent to normal, non-annotated speech by a generally enthusiastic speaker.

When characterizing a hot spot that contains varying levels of hotness, the highest level of hotness present is used to characterize the entire hot spot.

6. Degree

The degree of the hotness follows the level of hotness. Hot spots with a level of hotness of either warm or hot are followed by one of three possible degrees of hotness: <->, <0>, and <+>.

For hot spots with a warm or hot level of hotness, degree allows for further shading of the hotness so that the hotness is not restricted to discrete classes.

The degree of hotness provides further precision to the level of hotness and is defined on the scale below:

```
        -   0   +        -   0   +

        --------------------------------------------------
               W                    H
```
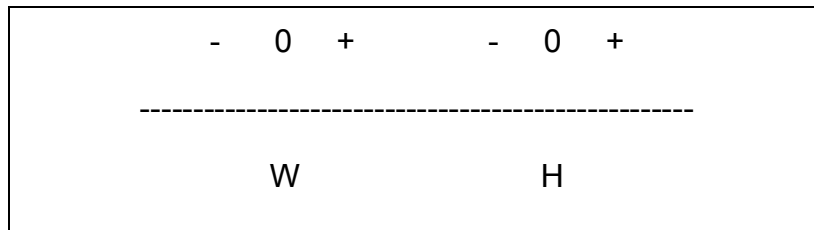
The <0> means that the hot spot is at the exact hotness that the class entails. The <-> and <+> mean that the hotness is slightly less or more, respectively, than what the hotness class alone can represent.

For hot spots with a lukewarm level of hotness, there is only degree possibility of <->. Therefore, a lukewarm level of hotness is restricted to the following two labels: <l> or <l->.

7. Type

A hot spot must be described by at least one hot spot type. When indicating the hot spot type, the letter that symbolizes the type (e.g., <a>, <d>, <o>, <c>, <g>, and <s>) is listed in the appropriate field within the label. If multiple types apply to a hot spot, those types which are present to the same and most salient degree are listed first and those that apply to lesser degrees are listed after the insertion of a semicolon. The primary type(s) may be listed in any order and the secondary types, if present, may be listed in any order also, so long as they are separated from the primary types with a semicolon.

The type of hot spot is determined by the nature of the involved utterances, not by the nature of the trigger, closure, or surrounding context outside of the hot spot. The types are explained below.

a -- Amusement

Amusement is marked whenever the speaker sounds amused. This usually, but not always, occurs when the speaker attempts to make a joke. It is important for the annotator to listen to precisely when laughter occurs. If the speaker laughs before or after talking, then, the speaker's actual utterances are not involved and would not indicate the need of marking a hot spot. Laughter and/or a signal of amusement by the speaker must occur *during* the speech in order for there to be

a hot spot. Laughter by other meeting participants are not indicative of an involved region of speech if the actual speaker shows no sign of amusement.

When a hot spot is labeled with the Amusement tag, the annotator may find that there was some other function to the utterance. However, if the main motivation for labeling a hot spot was because some level of amusement is heard, then amusement should be the tag before the semicolon. Other tags may be marked after the semicolon.

For more information on the Amusement tag, refer to the sections below that explain the Agreement and Suggestion tag usage.

c -- Clarification

Clarification is marked when a speaker either requests clarification by another meeting participant or when a speaker is clarifying a point for other meeting participants. Requests for clarification may come in the form of questions or in the form of statements that invite feedback from others. When a speaker provides clarification for others, it may be in the form of further elaboration on a point that the speaker has already said, or, it can be in the form of narrative answer to a question that another asked. In order for a region of utterances to be marked with a Clarification tag, the annotator must hear some sort of "heated" level of curiosity or involvement rather than simply spotting a clarification request or clarifying response.

d -- Disagreement

Disagreement is marked when a speaker has a stern tone or a frank manner of disagreeing. If the speaker is "sugar-coating" his or her disagreement so that the disagreement is said politely, it is unlikely that the speaker's utterances would constitute a hot spot.

g -- Agreement

Agreement is marked when a speaker shows a level of enthusiasm or greater emotional involvement in expressing agreement.

Sometimes, a speaker may express this greater involvement with laughter during his/her utterance that expresses agreement. In such cases, the reason such utterances constitute a hot spot is primarily because of Agreement rather than Amusement. A level of amusement arises *because* of the agreement as opposed to the agreement arising because of amusement. Therefore, laughter contained in utterances that indicate Agreement should be marked with a <g> *before* the semicolon. Depending on how much laughter there is determines whether an <a> tag is to be placed before or after the semicolon.

o -- Other

The Other tag is used to indicate any level of surprise, excitement, curiosity, or interest on the part of the speaker.

s -- Suggestion

The Suggestion tag is used when the speaker suggests something with a certain level of amusement, charisma, or enthusiasm. Do note, however, that if amusement is expressed during the speaker's utterances, then an <a> should be marked before the semicolon and an <s> should be marked after the semicolon.

8. Utterance Tag

Everything after the first period is part of the utterance tag. The order of the letters after the first period is inconsequential.

a: The <a> marks the trigger of a hot spot. A trigger is usually uninvolved but sometimes, triggers are also marked as involved utterances. If the trigger is warm, the utterance should be marked with an <a.b>.

n: The <n> tag marks the "opening" of a hot spot when there is no trigger. This will often occur when a whole chunk of utterances before hand initiates the beginning of a hot spot.

Here is an example of the use of <n> to mark the opening of a hot spot when there is no trigger. Previous utterances before the opening of the hot spot are included for context; the actual occurrence of <n> occurs in the last line of the extract.

Example 2: Bmr009

700.38,703.23,Bmr009-c1,s,,,,i think it's - to start with it's s- - your - your idea of simplifying .

703.34,707.69,Bmr009-c1,s.%-,,,,starting with something that you can see eh you know without the extra layers of ==

708.081,708.331,Bmr009-c3,fg,,,,right .

708.331,710.181,Bmr009-c3,s.%--,,,,cuz if energy doesn't matter there like ==

710.524,711.974,Bmr009-c3,s.%--,,,,i don't think this is true but what if ==

```
711.351,712.271,Bmr009-c4,s,4a,,,to study individual .

713.004,713.384,Bmr009-c3,qw^br,4b-2.5a-2,,,hmm ?

713.095,713.325,Bmr009-c1,s^fa,,,,sorry .

713.325,713.535,Bmr009-c1,qw^br,4b-1.5a-1,,,what ?

713.552,714.512,Bmr009-c4,s,5b.6a,,,to study individual .

715.04,717.16,Bmr009-c1,s^cs,6b-1.7a,,,well you - you - you don't
have to study everybody individually .

715.055,719.095,Bmr009-c3,s.%--,6b-2,,,well to study the simplest
case to get rid of extra ==

715.27,719.029,Bmr009-c4,s.%-,,,,the - the - the - but - consider ==

717.16,718.12,Bmr009-c1,s^e,6b-1+.7a+,,,but just simple case .

718.12,720.45,Bmr009-c1,s^e,,,,and the one that has the lot of data
associated with it .

720.017,720.317,Bmr009-c3,s^aa,7b,,,right .

720.936,723.396,Bmr009-c3,s.%--,,,,cuz what if it's the case and i
don't think this is true ==

722.62,724.25,Bmr009-c0,s^ba,,**10=w-=a.n.b.e**,,that was a great
overlap by the way .
```

b: A <b> marks an involved utterance or portion of speech. Involvement is speaker-dependent and only pertains to how an utterance is uttered as opposed to surrounding utterances or laughter (cf. `Level of Hotness'). While <b> marks a rather emotional involvement, the label <r> (see explanation of <r> tag below) refers to a raised voice that some speakers use in order to keep the floor independently of emotional involvement. The labels <b> and <r> can co-exist.

 b-: This tag was created for the sake of the annotators. A <b-> is marked when the annotator is not entirely sure whether the utterance at hand is involved or not. Ultimately, <b-> tags may be collapsed with the <b> tags.

r: This tag is used when any utterance, whether in a hot spot or not, is spoken with a raised voice. It is important to note that this tag is not used solely when people interrupt or when there is overlap. The <r> tag can occur independently of the involvement tag <b> (see explanation of <b> tag above).

r-: This tag is used when it is unclear for the annotator whether there is a raised voice. This tag may be collapsed with the <r> tags.

c: This tag denotes a closure. All hot spots must end with either a <c> or an <e>. A closure can be differentiated from an ending because it is more semantically related to the hot spot. A closure can be involved or uninvolved. Sometimes, a hot spot will consist of only one utterance which then serves as both, starter and finisher at the same time.

e: All hot spots must end with either a <c> or an <e>. Endings mark the last involved utterance. Therefore, <e> should always be accompanied by a <b>.

Possible Starters:

<a>, <a.b>, <n.b>

Possible single-utterance hot spots:

<a.b.e>, <n.b.e>, <a.b.c>, <n.b.c>

Possible Finishers:

<c>, <c.b>, <e.b>

9. Dividing Utterances (<**>)

If the first part of an utterance requires utterance tag(s) that are different than the latter part of the utterance, then double stars (<**>) maybe used. In such cases, it is necessary to mark double stars in both the hot spot and transcription fields of TableTrans.

For example, in the following two examples, the utterance is labeled with <b**c> to signify that the first part of the utterance is an involved portion and the latter part is an uninvolved closure.

> **Example 3: Bed015**
>
> 466.6,468.06,Bed015-c5,s,13a+,**b**c**,,so that- - ** that's why - that's why i'll stick with it .
>
> 2657.72,2660.2,Bed015-c5,s^aa|s^na,17b,**b**c**,,yeah | exactly ** exactly .

In the following example, the utterance is labeled with <a**b> to signify that the first part of the utterance is a cold trigger and the latter part is an involved region.

> **Example 4: Bmr009**
>
> 1059.9,1069.16,Bmr009-c4,s,10a,**60=l=a;c.a**b**,,i - i - i - i understood uh that eh i - i had to finish ****** by the moment with the and - and concentrate my - my energy in that problem .

In cases where an utterance tag (e.g. <b>) pertains to a whole utterance but another utterance tag (e.g. <r>) is only assigned to a portion of that utterance the use of <**> is necessary to separate the two portions of the utterance. The utterance should be marked with the utterance tag pertinent to the entire utterance on both sides of the double stars. The utterance tag pertinent to only one portion of the utterance should be marked on that side of the double stars.

For example, in the following utterance, the utterance is labeled with <r**b.r> to signify that the entire utterance requires an <r> but only the part before the double stars necessitate a <b>.

> **Example 5: Bmr009**
>
> 342.604,346.654,Bro014-c2,s,,**20=l=a.a.r**b.r**,FBH,so after i did that ****** then what i wanted to do was try increasing the number of mixtures .

10. Utterances with Marked and Unmarked Portions (<0>)

Two sets of double starts and a <0> are used in situations where the first part of an utterance needs one utterance tag, the middle part of the utterance should remain unmarked, and the last part of the utterance has yet another utterance tag.

> **Example 6:**
>
> **r**0**b** portion of utterance spoken with raised voice ****** unmarked portion ****** involved portion of utterance

Corresponding double stars should always be labeled in the hot spot field of TableTrans if double stars are inserted in the transcription field.

11. Two Labels for One Utterance (<##>)

A double pound sign (<##>) is used when two labels are required for one utterance. This usually occurs when an utterance is both the finisher of a previous hot spot and the starter of a new hot spot. The double pound sign is only marked in the label, not in the transcription field.

If an entire utterance is spoken with a raised voice but <##> are used in annotation, then an <r> should be marked on both sides of the <##>. The <##> are only marked in the hot spot column of TableTrans.

12. More Complex Tagging (<##> and <**>)

Sometimes there are situations where both double pound signs (<##>) and the double stars (<**>) are used. In the example below, the *whole* utterance functions as starter with the first portion being involved. At the same time the first portion serves as an involved ending for a previous hot spot.

Example 7: Bmr012

1767.17,1778.71,Bmr012-c2,qy^cs.%-,75a,**b.e**0##80=l-=go.a.b**a**,,is - is there any way ** we can have you know - like a - a wireless microphone that you pass around to the people who you know - the extra people for the times they wanna talk that ?==