Spotting Agreement and Disagreement:
A Survey of Nonverbal Audiovisual Cues and Tools

Konstantinos Bousmalis
Computing
Imperial College
London, UK
k.bousmalis@imperial.ac.uk

Marc Mehu
CISA
Uni. Geneva
Geneva, Switzerland
marc.mehu@unige.ch

Maja Pantic
Computing/EEMCS
Imperial College/Uni. Twente
London, UK/Enschede, NL
m.pantic@imperial.ac.uk

Abstract

While detecting and interpreting temporal patterns of non–verbal behavioral cues in a given context is a natural and often unconscious process for humans, it remains a rather difficult task for computer systems. Nevertheless, it is an important one to achieve if the goal is to realise a naturalistic communication between humans and machines. Machines that are able to sense social attitudes like agreement and disagreement and respond to them in a meaningful way are likely to be welcomed by users due to the more natural, efficient and human–centered interaction they are bound to experience. This paper surveys the nonverbal cues that could be present during agreement and disagreement behavioural displays and lists a number of tools that could be useful in detecting them, as well as a few publicly available databases that could be used to train these tools for analysis of spontaneous, audiovisual instances of agreement and disagreement.

1. Introduction

Agreements and disagreements occur daily in human–human interaction, and are inevitable in a variety of everyday situations. These could be as simple as finding a location to dine and as complex as discussing about notoriously controversial topics, like politics or religion. Agreement and disagreement are frequently expressed verbally, but the nonverbal behavioral cues that occur during these expressions play a crucial role in their interpretation [13]. That is naturally the case not only for agreement and disagreement, but for all facets of human social behavior, including politeness, flirting, social relations, and other social attitudes [78].

Machine analysis of nonverbal behavioral cues (*e.g.* blinks, smiles, nods, crossed arms, etc.), have recently been the focus of intensive research, as surveyed by Pantic *et al.* in [56, 58]. Similarly, significant advances have been made in the area of affect recognition (for exhaustive surveys, see [29, 82]). However, research efforts on the machine analysis of social attitudes are still at a rather early stage [56, 78].

There is no overview available, to the best of our knowledge, of nonverbal behavioral cues exhibited during agreement and disagreement. This paper attempts to fill this gap and to be the first step towards our eventual objective: creating a system that can automatically detect the relevant behavioral cues, and spot agreement or disagreement based on both their presence and temporal dynamics.

In this paper we list *(a)* different nonverbal behavioral cues relevant to detecting agreement and disagreement, *(b)* a number of tools that can detect these cues, and *(c)* a list of databases that can prove useful in the development of an automated system for (dis)agreement detection.

Note that we are interested only in those cues that can be detected using a monocular audiovisual data capture system. The main reason for this choice is the fact that the average user has a monocular camera connected to their computer system and hence, any output from this research will be directly applicable in standard user applications, without the need for additional and expensive equipment (such as biosensors, thermal cameras, *etc.*). Furthermore, it will be possible to directly apply the research findings for automatically analyzing and detecting agreement and disagreement in television data, such as televised political debates.

2. Agreement and Disagreement

Distinguishing between different kinds of agreement and disagreement is difficult, mainly because of the lack of a widely accepted definition of (dis)agreement [13]. Ekman [18] talked about listener's expressions of agreement and disagreement, distinguishing them from the relevant speaker's expressions. Argyle [1] specifically discussed the fact that speakers attend to listeners for nonverbal signals that not only serve as feedback to the process of the conver-

978-1-4244-4799-2/09/$25.00 ©2009 IEEE

sation, but also as an expression of the listener's opinion. Seiter et al. [67–69] have specifically discussed the importance of listener's expressions of disagreement.

Based on the findings reported by Poggi [62], we can distinguish among at least three ways one could express (dis)agreement with:

Direct Speaker's (Dis)Agreement: A speaker uses specific words that convey direct (dis)agreement, *e.g.* "I (dis)agree with what you have just said".

Indirect Speaker's (Dis)Agreement: A speaker does not explicitly state his or her (dis)agreement, but expresses an opinion that is congruent (agreement) or contradictory (disagreement) to an opinion that was expressed earlier in the conversation.

Nonverbal Listener's (Dis)Agreement: A listener expresses non-verbally her (dis)agreement to an opinion that was just expressed. This could be via auditory cues like "mm hmm" or visual cues like a head nod or a smile. (For a full list of the nonverbal cues that can be displayed during (dis)agreement, see Tables 1 and 2.)

Moreover, displays of agreement, and especially disagreement, can often be accompanied by expressions of emotions like anger, boredom, disgust or frustration as is the case for disagreement [27, 28, 68]. Hence, if the aim is to develop an automated system for (dis)agreement detection, automatic recognition of these affective states should be a part of the system as well.

In addition, Pomerantz [63, 64] describes disagreement as a dispreferred activity, and states that a weak agreement could actually be a preface to an act of disagreement. This makes the problem of (dis)agreement analysis truly complex. In this paper, we leave this aspect out of discussion.

3. Cues of Agreement and Disagreement

3.1. Cues of Agreement

Table 1 contains a list of all cues that can possibly be present during an agreement act. The most prevalent cue seems to be the **Head Nod** which is believed to be a nearly–universal indication of agreement [14, 50]. **Listener Smiles** are also rather indicative. However, both cues could have different meanings [7, 30], as further explained in Section 3.3.

When it comes to **Eyebrow Raise**, it is believed that it occurs in combination with other agreement–relevant cues particularly during an act of Nonverbal Listener's Agreement [18, 66]. Cohen [13] states that **Laughter** could also increase the reliability of any reasoning about detecting agreement, however there is no statistically grounded work on that, as far as we know. Finally, although **Sideways Leaning**, *e.g.*, leaning on a wall due to relaxation is referred to as an agreement cue by Bull [9] and reiterated by Argyle [1]. However, it is specifically discredited by Bull himself [10] as a weaker sign of agreement.

Human's communication system is fairly complex and it is unlikely that receivers will form intricate representations of attitude on the basis of a single cue. In fact, people most probably infer attitudes like agreement by using a combination of such cues, or through the perception of second order dynamic processes that involve these cues. For example **Mimicry** is a mutual imitation of the interlocutor's nonverbal behaviour and is believed to foster affiliation, agreement, and liking [12]. Mimicking the other person's positive behaviour such as nod or smile could therefore be interpreted as agreement; while the presence of the cue on its own might just signal something else, like submissiveness or interest.

3.2. Cues of Disagreement

When it comes to disagreement, it seems that a head shake is the most common cue. A **Head Shake** could specifically mean the refusal or reluctance to believe what is being said [18]. However, much like the head nod and the smile, this signal can have different purposes (look at Section 3.3 below).

Ironic smiles are a result of a conflict between two set of muscles and therefore are not as naturally occurring as benign smiles [1, 65]. Similar to the ironic smile is the **Cheek Crease**, during which a lip corner is pulled back strongly, deliberately distorting a smile to convey sarcasm [50]. These cues seem to be present in expressions of spontaneous and posed disagreement [50, 68].

Ekman [18] specifies that the **Eyebrow Raise**, or "scowling", as referred to by Seiter et al. [67], may indicate lack of understanding. However, it can also indicate, like the head shake, a listener's inability to believe what the speaker is saying or has just said. It can even express a "mock astonishment", when combined with a raised upper eyelid and/or a jaw drop.

Morris [50] mentions a number of disagreement–related cues. One of them is the **Nose Flare**, a result of the contraction of the muscles on either side of the nose, which is often accompanied by a sharp intake of air. Morris also mentions the **Head Roll**, which is the action of repeatedly tilting the head left and right expressing doubt. The **Sudden 'Cut Off'** is a gaze avoidance in which the head is turned fully away from the speaker. The **Leg Clamp**, though not specifically linked to disagreement, signifies stubbornness, as if the conversation participant was saying: "My ideas, like my body, are clamped firmly in position and will not budge an inch" [50]. The **Forefinger** and **Hand Wag**, during which an erect forefinger or a hand with the palm out-

CUE	KIND	REFERENCES
Head Nod	Head Gesture	[1, 14, 25, 30, 41, 50, 66]
Listener Smile/Lip Corner Pull (AU12, AU13)	Facial Action	[1, 7, 50]
Eyebrow Raise (AU1, AU2) + other agreement cues	Facial Action	[66]
AU1 + AU2 + Head Nod	Facial Action, Head Gesture	[16, 18]
AU1 + AU2 + Smile (AU12, AU13)	Facial Action	[16, 18]
AU1 + AU2 + Agreement Word	Facial Action, Verbal Cue	[16, 18]
Sideways Leaning	Body Posture	[1, 9, 30]
Laughter	Audiovisual Cue	[13]
Mimicry	Second–order Vocal and/or Gestural Cue	[1, 30, 35]

Table 1. Cues of Agreement. For relevant descriptions of AUs, see FACS [19].

wards, respectively, is wagged from side–to–side has a dissenting meaning. The **Neck Clamp**, the **Lip Bite** accompanied by a vigorous head shake, and the **Clenched Fist** signal anger with what is being said. The **Hand Cross** is simply a two–handed version of the hand wag. The **Hand Chop** is the action during which a hand imitates an axe, and the **Hand Scissor** is the action during which the hands imitate the blades of a pair of scissors. Morris mentions that both are often used unconsciously during a heated discussion. **Arm Folding** is widely known as signifying a defensive attitude and could also signify disagreement, *e.g.*, in situations where one participants is being verbally attacked in a strong disagreement [9, 25, 50].

Another very interesting cue is the **Throat Clearing**. Givens [25] states that disagreement and uncertainty can act like chemicals or food irritants and cause this signal. Givens also mentions that **Self–manipulation**, *e.g.*, a finger on the lips, massaging a hand, or a chin rub, can provide self–comfort when politeness prevents a listener from expressing disbelief and disagreement. Moreover, Givens argues that a sudden appearance of **Slightly Parted Lips** is a strong signal of nonverbal listener's disagreement. This is in agreement with Ekman's [18] finding that a listener's preparatory–to–speech mouth movement signals a desire to take the floor. Givens also considers a **Lip Pucker** to be the first sign of disagreement.

Disagreement could also be inferred by second order cues such as interruption, delay in responding, or utterance length. For example, Greatbach et al [28] argued that disagreement can be stronger if an **Interruption** and overlapping speech occur. Similarly, **Delays** in responding could be characteristics of a dispreferred activity, such as a disagreement act [63, 64]. In these two examples, it is not the act of speaking or not speaking *per se* that conveys disagreement but the act of violating implicit rules of turn-taking in a conversation. Note, however, that there are certain cases where disagreement becomes the preferred activity, as is the case with responses to compliments [53]. Finally, **Utterance Length** has been shown to be particularly longer in disagreement than in agreement acts [13, 24].

Table 2 shows a complete list of cues associated with disagreement.

3.3. Backchannel Signals: Nods, shakes and smiles

Ekman [18] states that although emotional expressions during conversations are a reaction to the "affective content", they can also relate to the participants' feelings regarding the nature and progress of the conversation itself, *i.e.*, they can serve as backchannel signals. Brunner [7] specifies that there are three levels of meaning a feedback backchannel could have, with the higher level implying and containing the lower ones. These are: **Level 1**—Involvement, **Level 2**—Level of understanding, **Level 3**—Actual response, *e.g.*, (dis)agreement.

Argyle [1] supports this by stating that backchannel signals may indicate attention and understanding, provide feedback like agreement, or be a part of mimicry, which in turn could signify agreement.

So, agreement and disagreement could be conveyed using backchannel signals and it could be argued that most of the implicit nonverbal cues of (dis)agreement are of this sort. For example, nods and shakes are two of the most common backchannel gestures. Nods usually have an affirmative meaning, especially if they're repeated and their amplitude is large. Smaller, one–way nods usually serve as signals of involvement in the conversation [1, 66]. However, it should be noted that head nods could also be negative [66]. Brunner [7] states that listener smiles can also be backchannels and are used in the same way as head nods. Brunner also argues that smiles act on the third level, *i.e.*, they provide a positive response to what is being said, they provide acknowledgment of understanding, and keep the listener involved in the conversation.

Head shakes are less common, and although they can have a dissenting meaning [1], they could also be part of a question or laughter [1, 30].

CUE	KIND	REFERENCES
Head Shake	Head Gesture	[1, 18, 30, 41, 50, 67, 69]
Head Roll	Head Gesture	[50]
Sudden 'cut off' (of they eye contact)	Head Gesture	[25]
Eye Roll	Facial Action	[41, 67–69]
Ironic Smile/Smirking [AU12 L/R (+AU14)]	Facial Action	[18, 67]
AU1 + AU2 + Raised Upper Lid (AU5)/...	Facial Action	[18]
.../Open Jaw Drop (AU26) with abrupt onset		
Barely noticeable lip–clenching (AU23, AU24)	Facial Action	[25]
Cheek Crease (AU14)	Facial Action	[50]
Lowered Eyebrow/Frowning (AU4)	Facial Action	[25, 69]
Lip Bite (AU32)	Facial Action	[50]
Lip Pucker (AU18)	Facial Action	[25]
Slightly Parted Lips (AU25)	Facial Action	[25]
Mouth Movement (Preparatory for Speech) (AU25/AU26)	Facial Action	[18]
Nose Flare (AU38)	Facial Action	[50]
Nose Twist (AU9 L/R and/or AU10 L/R and/or AU11 L/R)	Facial Action	[50]
Tongue Show (AU19)	Facial Action	[25]
Suddenly Narrowed/Slitted Eyes (fast AU7)	Facial Action	[25]
Arm Folding	Body Posture	[9, 25, 50]
Head/Chin Support on Hand	Body/Head Posture	[9, 25, 50]
Large Body Shift	Body Action	[25]
Leg Clamp (the crossed leg is clamped by the hands)	Body Posture	[50]
Sighing	Auditory Cue	[68]
Throat Clearing	Auditory Cue	[25]
Delays:Delayed Turn Initiation, Pauses, Filled Pauses	Second–order Auditory Cue	[13, 24, 28, 32, 63, 64]
Utterance Length	Second–order Auditory Cue	[13, 24]
Interruption	Second–order Auditory Cue	[28]
Clenched Fist	Hand Action	[25, 50]
Forefinger Raise	Hand Action	[50]
Forefinger Wag	Hand Action	[50]
Hand Chop	Hand Action	[50]
Hand Cross	Hand Action	[50]
Hand Wag	Hand Action	[50]
Hands Scissor	Hand Action	[50]
Neck Clamp	Hand/Head Action	[50]
Self–manipulation	Hand/Facial Action	[25, 50]
Head Scratch	Head/Hand Action	[50]
Gaze Aversion	Gaze	[66]

Table 2. Cues for Disagreement. For relevant descriptions of AUs, see FACS [19]

4. Detection Tools

Although in some cases detecting the cues in Tables 1 and 2 is rather straightforward, as is the case with cues that correspond to Action Units, there are cues that are known to be hard to detect. Two such examples are **Arm Folding** and **Head and Chin Support on a Hand**. [58]

However, there are known techniques that would be able to detect most of the cues listed in Tables 1 and 2. For example, most of the current head pose estimation computer–vision systems (for an exhaustive survey refer to [51]) can be adjusted for detection of **Head Nods and Shakes**, probably the most important cues for our objective. A system that can detect nods and shakes particularly well is the work of Morency *et al*. [47, 48].

There are a few attempts to automatically detect **Mimicry**, one of which is by Meservy *et al*. [45]. Keller *et al*. [35] also mention the possibility of using Motion Energy Analysis [6] to analyze the synchrony between the movements of the participants in a dyadic conversation. Pentland [59] measures mimicry (or "mirroring", as called in [59]) in conversational audio patterns, by using auditory backchannels and short words.

The hand and body actions of **Forefinger Wag**, **Hand Wag**, **Hand Cross** and **Hands Scissor** could be detected with adapted versions of human activity detection methods such as the work of Oikonomopoulos *et al*. [54], Marszałek *et al*. [42], Mikolajczyk *et al*. [46], Laptev *et al*. [37], Niebles *et al*. [52] and Shechtman *et al*. [70]. Actions like **Leg** or **Neck Clamp** and **Arm Folding** could also be detected with adaptions of these methods, but with more difficulty, and both dynamic and static features would have to be used for better results. Motion History Images [6] could also be used for such actions, but they have proven to be particularly sensitive to, *e.g*., different clothing. The latter actions could also be detected by the arm and hand tracker of Buehler *et al*. [8]. Most of the other hand actions, and especially **Hand Chop**, **Hands Scissor**, **Hand Wag and Cross** could also be detected by adapting the latter work. **Clenched Fist** and **Forefinger Raise and Wag** seem to be able to be detected by adapting the hand gesture interface system implemented by Ike *et al*. [33]. Most of the aforementioned hand gestures and some self–manipulation gestures like face/lips touching can be detected by sign language recognition methods such as that by Ding and Martinez [15].

When it comes to automatically detecting facial actions, significant advances have been made over the past ten years. Table 4 lists examples of the state–of–the–art systems, omitting older ones that cannot detect Action Units (AUs) in combinations, as discussed and surveyed by Tian *et al*. [72]. AUs are atomic facial signals, the smallest visually discernible facial movements. FACS [19] defines 9 upper face AUs, 18 lower face AUs, and 5 miscellaneous AUs. The most comprehensive works in automatic AU detection are those of Koelstra and Pantic [36] and Vural *et al*. [79], as they detect most of the AUs defined in FACS [19], including those that could be cues of (dis)agreement. The former also enables analysis of temporal dynamics of AUs, which could prove very important when distinguishing, for example, a smile(slow symmetric action) from a smirk (fast asymmetric action). However, these methods will not work particularly well if rigid head movements are not properly dealt with, which is usually a problem with naturalistic, spontaneous data. The work of Valstar and Pantic [76] can also detect many of the AUs listed in tables 1 and 2, including their temporal dynamics, while handling problems with head movement registration rather well. For exhaustive surveys on the topic, see Pantic *et al*. [55, 58].

Smiles relate to AU12 and AU13, which can be recognized by many AU detection systems, as one can see in Table 4. However, the work done by Valstar *et al*. [75] is

CUE	REFERENCES
Head Nod/Shake	[20, 22, 34, 47, 71]
Mimicry	[35, 45, 59]
Smiles vs Smirks	[75]
Utterance Length	[32]
Laughter	[60, 61, 74]
Eye Roll	[20]
Head Roll	[20]
Filled Pause	[2, 23, 26, 80]
Pause	[4, 43]
Interruption	[38, 40]
Throat Clearing	[44]
Tongue	[21, 83]
Sudden 'Cut Off'	[3]
Hand Scissor/Wag/Cross	[8, 37, 42, 46, 52, 54, 70]
Clenched Fist/Forefinger Raise	[33]
Forefinger Wag	[33, 37, 42, 46, 52, 54, 70]

Table 3. Tools for detecting cues for agreement and disagreement

able to distinguish between spontaneous and posed smiles, which could prove particularly useful in differentiating between genuine, benign smiles and ironic ones (*e.g*., smirks).

Sudden 'Cut Off' can be detected by adapting methods aimed at detecting the focus of one's attention such as the recent work of Ba and Odobez [3]. Other works on head tracking [51] and on gaze tracking [49] can be adapted for this purpose as well. Recent work can also detect **Laughter** and distinguish it from speech, using auditory cues [74] or a fusion of auditory and visual cues [60, 61]. Finally, the work of Matos *et al*. in [44] can detect **Throat Clearing** as a sub–goal to cough detection.

Tables 3 and 4 list some of the discussed, recently proposed tools that could be used/adapted to detect the cues relevant to agreement and disagreement, as those listed in Tables 1 and 2. Yet, in spite of this obvious progress in automatic analysis of various behavioural cues, no effort has been reported so far towards automatic analysis of (dis)agreement in naturalistic data. The only work in the field is that by el Kaliouby and Robinson [20], which attempted (dis)agreement classification of acted behavioural displays based on head and facial movements. Detection of these signals in naturalistic data is yet to be attempted.

5. Databases of Relevant Naturalistic Data

To develop and evaluate automatic analyzers capable of dealing with naturalistic occurrences of agreement and disagreement as defined earlier in this paper, large collections of training and test data, recorded in naturalistic settings, are needed.

Televised political debates provide an interesting platform for analyzing agreement and disagreement–related

System	AUs Detected																	
	1	2	4	5	9	10	11	12	13	14	18	19	23	24	25	26	32	38
Tian *et al*. (2001) [72]	√	√	√	√	√	√		√						√	√			
el Kaliouby *et al*. (2005) [20]	√	√						√	√	√	√				√	√		
Pantic *et al*. (2005) [57]	√	√	√	√		√		√	√		√		√	√	√	√		√
Bartlett *et al*. (2006) [5]	√	√	√	√	√	√	√	√		√			√	√	√	√		
Littlewort *et al*. (2006) [39]	√	√	√	√	√													
Yang *et al*. (2007) [81]	√	√		√		√		√		√								
Valstar *et al*. (2007) [76]	√	√	√	√	√	√		√	√		√			√	√	√		
Koelstra *et al*. (2008) [36]	√	√	√	√	√	√	√	√	√	√	√		√	√	√	√		
Vural *et al*. (2008) [79]	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Tong *et al*. (2009) [73]	√	√	√				√						√	√	√	√		

Table 4. AU detection systems

cues. Since the first televised political debates of the 1960's, debates have become more common, and the audience actually expects the participation of political figures in them. [68] At the same time, the presentation of such debates has evolved from a single–screen approach to multiple split screens, where every reaction each participant makes is available for examination, regardless of who the speaker is. [67] Even if only a single screen is used, the director of the debate will often use close–ups of the speaker or the listeners to give access to the nonverbal aspect of their behavior. [31] Research has suggested that those watching the debates perceive as less likable the participants who attempt to belittle a debate opponent via cues of nonverbal listener's disagreement. Interestingly enough, political figures are still prepped to display certain cues for that purpose, and hence this is an interesting case of acted agreement and disagreement.

Canal9[1] [77] is an example of a database of political debates. The database contains a total of over 42 hours of real televised debates on Canal 9, a Swiss television network. There is always a moderator and two sides that argue, with one or more participants on each side. Although this is a "political" debates database, the subjects are not always politicians, and the public opinion does not matter as much. Hence, instances of masked or acted (dis)agreement mentioned above, are rare. The debates are pre–edited in one feed and more than one camera angles are used.

Roma Tre Political Debates[1] is another such database. It contains ten political talk shows and pre–election debates aired on Italian television networks. The number of participants ranges from two to six and each video lasts from 60 to 90 minutes.

The *Green Persuasive Dataset*[1] is a database of 8 recorded instances of attempts by strong pro–green individuals to convince others to adopt a 'greener' lifestyle. There are many instances of agreement and disagreement. Each discussion is a dyadic interaction and lasts from 25 to 48 minutes.

Other databases that could be useful for training and testing automated tools for (dis)agreement detection would be those capturing the instances of human–human or human–computer interaction, in which occurrences of (dis)agreement are very common. Such databases are group meetings recordings like the *AMI Dataset*[1] [11] and human–virtual character interaction recordings like the *SAL Dataset*[1] [17]. For an exhaustive overview of such databases, see [29, 82].

6. Conclusion

This paper has attempted to provide an overview of the cues useful for detecting agreement and disagreement. It has also attempted to provide a list of the state–of–the–art tools that can be used/adapted to detect these cues. Finally, a list of databases that could be used to train and test automated tools for (dis)agreement detection is also provided. Hence, we hope that the paper can serve as an introductory reading to al researchers interested in the problem of automatic detection of agreement and disagreement.

Acknowledgments

This work has been funded in part by the EC's 7th Framework Programme [FP7/2007-2013] under the grant agreement no 231287 (SSPNet). The work of Konstantinos Bousmalis is also funded in part by the EC's 7th Framework Programme [FP7/2007-2013] under grant agreement no 211486 (SEMAINE). The work of Maja Pantic is also funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

[1]These databases and more information about them can be found online at the SSPNet web portal (http://www.sspnet.eu).

References

[1] M. Argyle. *Bodily Communication*. 2 edn., 1988. Chapter 7.

[2] K. Audhkhasi, K. Kandhway, O. Deshmukh, and A. Verma. Formant–based technique for automatic filled-pause detection in spoken english. *Proc. IEEE Int'l Conf. Acoustics, Speech and Signal Processing*, pp. 4857–4860. 2009.

[3] S. Ba and J. Odobez. Recognizing Visual Focus of Attention from Head Pose in Natural Meetings. *IEEE Trans. Systems, Man, and Cybernetics, Part B*, 39(1):16–33, 2009.

[4] D. Baron, E. Shriberg, and A. Stolcke. Automatic punctuation and disfluency detection in multi–party meetings using prosodic and lexical cues. *Proc. Int'l Conf. Spoken Language Processing*, pp. 949–952. 2002.

[5] M. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan. Fully automatic facial action recognition in spontaneous behavior. *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 223–230. 2006.

[6] A. Bobick and J. Davis. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(3):257–267, 2001.

[7] L. J. Brunner. Smiles can be back channels. *Journal of Personality and Social Psychology*, 37(5):728–734, 1979.

[8] P. Buehler, M. Everingham, D. Huttenlocher, and A. Zisserman. Long term arm and hand tracking for continuous sign language TV broadcasts. *Proc. Conf. British Machine Vision*, pp. 1105–1114. 2008.

[9] P. Bull. *Posture and Gesture*, chap. 5: The Encoding of Disagreement and Agreement, pp. 62–69. Pergamon Press, 1987.

[10] P. Bull. *Posture and Gesture*, chap. 6: The Decoding of Interest/Boredom and Disagreement/Agreement, pp. 70–84. Pergamon Press, 1987.

[11] J. Carletta. Unleashing the killer corpus: experiences in creating the multi–everything AMI Meeting Corpus. *Language Resources and Evaluation Journal*, 41(2):181–190, 2007.

[12] T. Chartrand and J. Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910, 1999.

[13] S. Cohen. A computerized scale for monitoring levels of agreement during a conversation. *University of Pennsylvania Working Papers in Linguistics*, 8(1):57–70, 2003.

[14] C. Darwin. *The expression of emotions in man and animals*. Oxford University Press, USA, 2002.

[15] L. Ding and A. Martinez. Modelling and recognition of the linguistic components in american sign language. *Image and Vision Computing Journal*, 27(12), 2009.

[16] A. Dittmann. Develpmental factors in conversational behavior. *Journal of Communication*, 22(4):404–423, 1972.

[17] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. McRorie, J. Martin, L. Devillers, S. Abrilian, A. Batliner, *et al.* The HUMAINE database: Addressing the collection and annotation of naturalistic and induced emotional data. *Lecture Notes in Computer Science*, 4738:483–500, 2007.

[18] P. Ekman. *Human Ethology*, chap. About Brows: Emotional and Conversational Signals. Cambridge Univ. Press, 1979.

[19] P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system. Salt Lake City: Research Nexus, 2002.

[20] R. el Kaliouby and P. Robinson. Real–time inference of complex mental states from facial expressions and head gestures. *Proc. IEEE Int'l Conf. Computer Vision & Pattern Recognition*, vol. 3, p. 154. 2004.

[21] Z. Fu, W. Li, X. Li, F. Li, and Y. Wang. Automatic tongue location and segmentation. *Proc. Int'l Conf. Audio, Language and Image Processing*, pp. 1050–1055. 2008.

[22] S. Fujie, Y. Ejiri, K. Nakajima, Y. Matsusaka, and T. Kobayashi. A conversation robot using head gesture recognition as para-linguistic information. *Proc. IEEE Int'l Workshop Robot and Human Interactive Communication*, pp. 159–164. 2004.

[23] M. Gabrea and D. O'Shaughnessy. Detection of filled pauses in spontaneous conversational speech. *Proc. Int'l Conf. Spoken Language Processing*, vol. 3, pp. 678–681. 2000.

[24] M. Galley, K. McKeown, J. Hirschberg, and E. Shriberg. Identifying agreement and disagreement in conversational speech: use of bayesian networks to model pragmatic dependencies. *Proc. Meeting Association for Computational Linguistics*, pp. 669–676. 2004.

[25] D. B. Givens. The nonverbal dictionary of gestures, signs and body language cue. Center for Nonverbal Studies Press, Sokane, WA, 2002.

[26] M. Goto, K. Itou, and S. Hayamizu. A real–time filled pause detection system for spontaneous speech recognition. *Proc. European Conf. Speech Communication and Technology*, pp. 227–230. 1999.

[27] J. Gottman, H. Markman, and C. Notarius. The topography of marital conflict: A sequential analysis of verbal and nonverbal behavior. *Journal of Marriage and the Family*, 39(3):461–477, 1977.

[28] D. Greatbatch. *Talk at Work*, chap. 9: On the management of disagreement between news interviewees. Cambridge University Press, 1992.

[29] H. Gunes and M. Pantic. Automatic, dimensional and continuous emotion recognition. *Int'l Journal of Synthetic Emotion*, 1(1), 2009.

[30] U. Hadar, T. Steiner, and F. C. Rose. Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, 9(4):214–228, 1985.

[31] F. Haumer and W. Donsbach. The rivalry of nonverbal cues on the perception of politicians by television viewers. *Journal of Broadcasting and Electronic Media*, 53(2):262–279, 2009.

[32] D. Hillard, M. Ostendorf, and E. Shriberg. Detection of agreement vs. disagreement in meetings: training with unlabeled data. *Proc. Conf. North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pp. 34–36. 2003.

[33] T. Ike, N. Kishikawa, and B. Stenger. A Real-Time Hand Gesture Interface Implemented on a Multi-Core Processor. *Proc. IAPR Conf. Machine Vision Applications*, pp. 9–12. 2007.

[34] S. Kawato and J. Ohya. Real-time detection of nodding and head-shaking by directly detecting and tracking the between-eyes. *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 40–45. 2000.

[35] E. Keller and W. Tschacher. Prosodic and gestural expression of interactional agreement. *Lecture Notes in Computer Science*, 4775:85–98, 2007.

[36] S. Koelstra and M. Pantic. Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics. *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition*. 2008.

[37] I. Laptev and P. Perez. Retrieving actions in movies. *Proc. Int'l Conf. Computer Vision*, pp. 1–8. 2007.

[38] C.-C. Lee, S. Lee, and S. Narayanan. An analysis of multimodal cues of interruption in dyadic spoken interactions. *Proc. European Conf. Speech Communication and Technology*, pp. 1678–1681. 2008.

[39] G. Littlewort, M. S. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. *Image and Vision Computing*, 24(6):615–625, 2006.

[40] Y. Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf, and M. Harper. Enriching speech recognition with automatic detection of sentence boundaries and disfluencies. *IEEE Trans. Audio, Speech, and Language Processing*, 14(5):1526–1540, 2006.

[41] V. Manusov and A. R. Trees. "Are You Kidding Me?": The Role of Nonverbal Cues in the Verbal Accounting Process. *The Journal of Communication*, 52(3):640–656, 2002.

[42] M. Marszałek, I. Laptev, and C. Schmid. Actions in context. *Proc. IEEE Conf. Computer Vision & Pattern Recognition*. 2009.

[43] M. Marzinzik and B. Kollmeier. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Trans. Speech and Audio Processing*, 10(2):109–118, 2002.

[44] S. Matos, S. Birring, I. Pavord, and H. Evans. Detection of cough signals in continuous audio recordings using hidden markov models. *IEEE Trans. Biomedical Engineering*, 53(6):1078–1083, 2006.

[45] T. Meservy, M. Jensen, J. Kruse, J. Burgoon, J. Nunamaker, D. Twitchell, G. Tsechpenakis, and D. Metaxas. Deception detection through automatic, unobtrusive analysis of nonverbal behavior. *IEEE Intelligent Systems*, 20(5):36–43, 2005.

[46] K. Mikolajczyk and H. Uemura. Action recognition with motion-appearance vocabulary forest. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8. 2008.

[47] L.-P. Morency, C. Sidner, C. Lee, and T. Darrell. Contextual recognition of head gestures. *Proc. ACM Int'l Conf. Multimodal Interfaces*, pp. 18–24. 2005.

[48] L.-P. Morency, C. Sidner, C. Lee, and T. Darrell. Head gestures for perceptual interfaces: The role of context in improving recognition. *Artificial Intelligence*, 171(8-9):568–585, 2007.

[49] C. Morimoto and M. Mimica. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding*, 98(1):4–24, 2005.

[50] D. Morris. *Bodytalk: A world guide to gestures*. Jonathan Cape, 1994.

[51] E. Murphy-Chutorian and M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(4):607–626, 2009.

[52] J. Niebles and L. Fei-Fei. A hierarchical model of shape and appearance for human action classification. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8. 2007.

[53] R. Ogden. Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38(10):1752–1775, 2006.

[54] A. Oikonomopoulos, M. Pantic, and I. Patras. Sparse B-spline polynomial descriptors for human activity recognition. *Image and Vision Computing*, 27(12), 2009.

[55] M. Pantic. Machine analysis of facial behaviour: Naturalistic and dynamic behaviour. *Philosophical Transactions of Royal Society B*, 2009.

[56] M. Pantic, A. Nijholt, A. Pentland, and T. S. Huang. Human–Centred Intelligent Human–Computer Interaction (HCI^2): How far are we from attaining it? *Journal of Autonomous and Adaptive Communications Systems*, 1(2):168–187, 2008.

[57] M. Pantic and I. Patras. Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. *Proc. IEEE Int'l Conf. Systems, Man and Cybernetics*, vol. 4, pp. 3358–3363. 2005.

[58] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang. Human computing and machine understanding of human behavior: A survey. *Lecture Notes in Computer Science*, 4451:47–71, 2007.

[59] A. Pentland. Socially aware, computation and communication. *Computer*, 38(3):33–40, 2005.

[60] S. Petridis and M. Pantic. Audiovisual laughter detection based on temporal features. *Proc. ACM Int'l Conf. Multimodal Interfaces*, pp. 37–44. 2008.

[61] S. Petridis and M. Pantic. Fusion of audio and visual cues for laughter detection. *Proc. ACM Int'l Conf. Content–based Image and Video Retrieval*, pp. 329–338. 2008.

[62] I. Poggi. *Mind, hands, face and body: Goal and belief view of multimodal communication*. Weidler, 2007.

[63] A. M. Pomerantz. *Second Assessments: A Study of Some Features of Agreements/Disagreements*. General sociology, University of California, Irvine, 1975.

[64] A. M. Pomerantz. *Structures of Social Action: Studies in Conversation Analysis*, chap. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. Studies in Emotion and Social Interaction. Cambridge University Press, 1984.

[65] W. Rinn. The neuropsychology of facial expression: a review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin*, 95(1):52–77, 1984.

[66] H. M. Rosenfeld and M. Hancks. *The Relationship of Verbal and Nonverbal Communication*, chap. The Nonverbal Context of Verbal Listener Responses, pp. 193–206. Walter de Gruyter, 1980.

[67] J. Seiter. Does communicating nonverbal disagreement during an opponent's speech affect the credibility of the debater in the background? *Psychological Reports*, 84:855–861, 1999.

[68] J. S. Seiter, H. J. Kinzer, and H. Weger. Background behavior in live debates: The effects of the implicit ad hominem fallacy. *Communication Reports*, 19(1):57–69, 2006.

[69] J. S. Seiter and H. Weger. Audience perceptions of candidates' appropriateness as a function of nonverbal behaviors displayed during televised political debates. *The Journal of Social Psychology*, 145(2):225–236, 2005.

[70] E. Shechtman and M. Irani. Matching local self-similarities across images and videos. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8. 2007.

[71] W. Tan and G. Rong. A real-time head nod and shake detector using HMMs. *Expert Systems with Applications*, 25(3):461 – 466, 2003.

[72] Y.-I. Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(2):97–115, 2001.

[73] Y. Tong, W. Liao, and Q. Ji. *Affective Information Processing*, chap. 10: Automatic Facial Action Unit Recognition by Modeling Their Semantic and Dynamic Relationships, pp. 159–180. Springer London, 2009.

[74] K. P. Truong and D. A. van Leeuwen. Automatic discrimination between laughter and speech. *Speech Communication*, 49(2):144 – 158, 2007.

[75] M. F. Valstar, H. Gunes, and M. Pantic. How to distinguish posed from spontaneous smiles using geometric features. *Proc. ACM Int'l Conf. Multimodal Interfaces*, pp. 38–45. 2007.

[76] M. F. Valstar and M. Pantic. Combined support vector machines and hidden markov models for modeling facial action temporal dynamics. *Lecture Notes in Computer Science*, 4796:118–127, 2007.

[77] A. Vinciarelli, A. Dielmann, S. Favre, and H. Salamin. Canal9: A database of political debates for analysis of social interactions. *Proc. IEEE Int'l Conf. Affective Computing and Intelligent Interfaces*, vol. 2. 2009.

[78] A. Vinciarelli, M. Pantic, and H. Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12), 2009.

[79] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan. Automated drowsiness detection for improved driving safety. *Proc. Int'l Conf. Automotive Technologies*. 2008.

[80] C.-H. Wu and G.-L. Yan. Acoustic feature analysis and discriminative modeling of filled pauses for spontaneous speech recognition. *The Journal of VLSI Signal Processing*, 36(2–3):91–104, 2004.

[81] P. Yang, Q. Liu, and D. Metaxas. Boosting coded dynamic features for facial action units and facial expression recognition. *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 1–6. 2007.

[82] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 31(1):39–58, 2009.

[83] L. Zhi, J.-Q. Yan, T. Zhou, and Q.-L. Tang. Tongue Shape Detection Based on B-Spline. *Proc. Int'l Conf. Machine Learning and Cybernetics*, pp. 3829–3832. 2006.