Backchannels must be seen
Arie Timmerman
Universiteit Twente
Op de Wal 4
7511NS, Enschede
The Netherlands
a.timmerman@student.utwente.nl

ABSTRACT
Backchannel-continuers occur in many places during a conversation. A lot of factors influence the non-verbal backchannel behavior. One of these factors is eye gaze. Eye gaze is where this research focuses on.

Since most backchannel-continuers are non-verbal, eye gaze is the most important tool to recognize backchannels. Recognizing backchannels is important for the speakers behavior. Otherwise he may reformulate his sentences without reason or clarifies himself unnecessarily. This research shows the importance of recognizing backchannels. Listeners do give most backchannels during gaze. This is needed for giving a speaker the opportunity to adapt to his listeners and to make sure the listeners motivation to give backchannel-continuers does not disappear.

Keywords
Backchannel-continuers, eye gaze, gaze shifts, communication, backchannel prediction model

1. INTRODUCTION
In conversations, people repeatedly look to each other which happens for multiple reasons. For example for seeking information or showing interest[1]. Less well known is the relation between eye gaze and the backchannel behavior.

Backchannel-continuers (often referred to as just backchannels) are signals of the listener to the speaker that he should continue with his conversation. These signals include head nods or simple utterances like yes, mm-hmm and so on. Without these signals a speaker may conclude he is not understood and reformulate his words.

What does eye contact has to do with this? This research focuses on the relation between eye movement of the speaker and the backchannel behavior of the listener. We already know that such a relation exists [14]. This paper further explores this relationship. Not only a direct relation between eye gaze shifts and backchannel occurrences is described. Also described is the relation between the way a speaker handles the backchannels he receives and how the listeners do react on this.

The communication between human and machine is in the last decades an upcoming topic. What is the best, and most natural way a machine can communicate with us. In order to model a virtual agent it is required to know things about communication between human beings. The results of the research are important to learn more about this topic. This may be useful to improve backchannel prediction models; models which use often multiple features of communication to determine possible positions for backchannel continuer as good as possible. Such models could be used in a virtual agent and in human-machine-communication systems in general.

1.1 Research Overview
Besides the relation between eye gaze and backchanneling this research describes in what manner the backchanneling influences the speaker. Does the reaction of a speaker influence the behavior of the listener? In other words, does the speaker give feedback on a backchannel which is import for the further backchanneling behavior? A listener does respond with a reason. He expects a response from the listener. Possible responses of a speaker as a reaction to backchanneling are for example to do nothing instead of reformulating his words or to quickly jump over to the next point when the listeners shows he understands what is said.

The exact reaction of a speaker on the backchanneling will not be studied. Will be studied is whether the listener changes his behavior when he does not get the expected response from the speaker.

How long does it take to trigger the backchannel after the speaker gazed at the listener and does the speakers response to backchannels influence the listeners backchannel behavior? At first this question must be answered. The first part of this question is answered before in older researches but has to be answered again to validate our data and to ensure the second part can be answered properly.

2. Related Work
Backchannel-continuers appear at certain position in conversation. For example when the speaker pauses for a moment, at the end of a sentence [3], after certain spoken words or after changing prosody features [12]. Besides these features, eye gaze is also an indicator for backchanneling. Rosenfeld and Hancks [14] showed many factors that influence the backchannel behavior. One of the factors they mention is gaze. If a speaker starts watching towards the listeners, that is he makes a gaze shift, the chance he backchannels increases a lot.

Rosenfeld and Hancks were not alone in this conclusion. Others came to the same and concluded that eye gaze shifts are good for predicting backchannels. Recently Morency et al. [11] have written a paper about a probabilistic multimodal to predict listeners backchannels. In this model they used features like prosody, spoken words and eye gaze.

It is not strange eye gaze shifts may trigger backchannels. It is what we expect when we think about it and is something you notice in real life. Besides that we know that listeners do watch more to a speaker then a speaker does to him [6]. A listener
normally sees the speaker most of the time and thus sees him shift his gaze.

Eye gaze interaction has an important role when people talk, for information receiving as well as sending being the first function the most important one [1]. With our eyes we express ourselves and also have the ability to control the turn taking process [11]. Studies have also made clear that people make more eye contact when they have positive feelings including liking the listener [4]. The papers mentioned before have shown that eyes do also have an important function when it comes to backchannel behavior. Making eye contact is likely to trigger a backchannel [11]. It is like we ask for a backchannel when looking to someone while having a conversation.

Ward and Tsukahara [15] were concerned with predicting backchanneling using prosodic features. They found out that 110 milliseconds of low pitch by itself is a fairly good predictor. Other factors like utterance end, rising intonation and specific lexical items account for less than one would expect. They also found out that speaker-produced cues can account for about half of the occurrences of backchannels. A speaker seems to have a lot of influence in the backchannel behavior.

All these studies wrote about when backchanneling occurs, what it triggers. Less about the importance of a speaker to react on listeners backchannel signals.

Also worth mentioning is that individual differences in eye contact are very large [2]. Differences are found for example between man and women. Some studies stated that the backchanneling behavior in general differs per culture [10]. For example differences are found between the Japanese and American culture. Since this is the case this research solely focuses on dutch speaking people.

In this research we must not forget that there are other verbal and non-verbal signals which may trigger backchannels as well. Gaze shifts only account for triggering a relatively small amount of all backchannels.

3. Methodology

For the analyzing the MultiLis Corpus [7] is used. This corpus was not specially designed for this research but it fits the needs of this research extremely well.

3.1 The corpus

The MultiLis Corpus is a set of 32 conversations. Each with one speaker and 3 listeners who all were unaware of the other listeners. The speaker talks to the listeners using a monitor and a camera. The camera is placed directly behind the monitor so it is possible to make eye contact.

The speaker only sees one of the listeners and is not aware of the other 2 listeners. All listeners see and hear the speaker as if they were having a one-to-one conversation.

Since the interaction between participants during the conversation is limited, it is hard for the participant to notice the actual setup of the experiment. Listeners are not allowed to ask for an explicit reaction of the speaker and thus are for example not allowed to talk, except for saying things like yes, mm-hmm etc. In total 8x4=32 conversations are made. One of these conversations is excluded from this research because one of the camera’s pointed at the listeners was broken. In another conversation the headphone of one of the listeners did not work but the videocamera did. The results of this conversation are used since verbal backchannels happen very rarely and this experiment is not focussed on this type of backchannels.

In total 29 men and 3 woman took part.

Figure 1 shows a schematic view of the experiment setting. Spk stands for speaker, eav for eavesdropper. In the illustration lis stands for listener. Both the eavesdroppers and the listener do see the speaker on a videoscreen and think the speaker sees him the same way. This only accounts for the listener. The eavesdroppers, as well as the listeners and the speaker are not told about this. This makes the setting as natural as possible and also gives the ability to study three different persons who backchannel differently to the same speaker as they were having a one-to-one conversation.

Figure 1: Experiment Setting.

3.2 Annotation

The videos, with a total length of 133 minutes, all have been annotated. All annotation data used in this research is gathered manually.

For this research it was needed to annotate the moment a speaker starts looking in the camera and the moment a listener reacts with a backchannel. In total 2767 backchannels made by the listener and 3389 gaze periods made by the speaker where found in 31 conversations.

A gaze period started at the video frame a speaker clearly looked into the camera and ended at the frame he clearly looked away. This also includes eye blinking.

A backchannel moment started at the frame a listener clearly started reacting to the speaker.

For every experiment we have two eavesdroppers files, one listener file and one speaker file. In this research sometimes a difference between the eavesdroppers and the listener files is made but normally it will be used as they were having the same type of data.

4. Analysing

4.1 More Gaze implies more backchannels

Earlier research already found that a gaze may trigger a backchannel. This research agrees with these previous works. Our data shows that the more a speaker starts gazing, the more likely backchannels occur.

As mentioned before a total of 3389 gaze periods were found. These gaze periods are much longer then the non-gaze periods. In fact over 3 quarters of all moments a speaker does not gaze, last only 1 second or less. In the corpus ca. 30% of the time a speaker gazes towards the listener. Around 75% of all backchannels start while the speaker is gazing which can not be by accident. Normally a listeners gazes about 2.5 to 3 times as many as the speaker does [1] and thus has the opportunity to view most gaze shifts a speaker makes.
The length of a gaze is far less important than the amount of gaze-moments. A direct relation between the length of the gaze and the amount or duration of the backchannels is not found in this research. What is found is a clear relation between the amount of moments a speaker starts gazing and the amount of backchannels as we see in Figure 2. In this graph we see on the vertical axis the number of gaze shifts per minute. On the horizontal axis we have the number of backchannels per minute. We see the relative amount of gazes increase when the relative amount of backchannel does.

Of course the graph does not show a straight line. Not everyone who faces more gaze shifts will respond with more backchannels. Individual differences could explain the peaks in the graph. It is also possible some speakers do watch with more or less expression which may trigger backchannels easier.

The peaks probably also happen because other factors play a part. For example it could be possible that a speaker who makes lots of gaze shifts talks very monotonous or does not take many pause moments which influences the backchanneling behavior.

It has happened that someone who faces lots of gaze shifts does not respond much. The opposite did not happen. Participants who respond all lot, the high peaks, all faces an average of at least 8 backchannel per minute.

This graph can not be used to show how important gaze is for backchanneling. It only shows that there is a clear relation between the amount of gaze shifts and the amount of backchannels.

**Figure 2: Gaze shifts per minute as a function of the amount of backchannels**

A relation between the total gaze-length and the amount of backchannels is not found.

These results confirm what earlier papers have written. Our experiment shows the same relation between gaze and backchannels.

### 4.2 The chance a backchannel occurs

The previous section made clear that a relation between gaze shifts and backchannels exists. Such a relation is nice but it is not clear what this relation exactly means. Do gaze shifts trigger backchannels directly or is the relation based on something else?

Intuitively we feel it is very likely a listener will react when we starts gazing towards him. But sometimes this happens and sometimes it does not. Since a speaker makes a lot of gaze shifts it is almost impossible for a listener to react on every moment the speaker starts gazing. In this research we show that still a good chance a backchannel occurs immediately after gaze shift exists.

Table 1 shows the average chance a listener will react after a gaze shift within a certain amount of time. The differences between the eavesdroppers and the listener immediately strike. The chance a listener reacts with a backchannel after gaze is always bigger than the chance a eavesdropper does. Although these differences seems to be small it is relatively seen pretty often.

**Table 1: The average chance a listener will react after a gaze shift within a certain amount of time**

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Eavesdroppers</th>
<th>Listeners</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 100 ms</td>
<td>0.008</td>
<td>0.011</td>
<td>0.009</td>
</tr>
<tr>
<td>&lt; 500 ms</td>
<td>0.045</td>
<td>0.060</td>
<td>0.050</td>
</tr>
<tr>
<td>&lt; 1000ms</td>
<td>0.106</td>
<td>0.129</td>
<td>0.114</td>
</tr>
<tr>
<td>&lt; 2500ms</td>
<td>0.292</td>
<td>0.358</td>
<td>0.314</td>
</tr>
</tbody>
</table>

The interval is chosen with reason. The human reaction time is in this research considered to be 100ms. The exact reaction time is unknown for this specific situation but in similar cases [5] this is found out to be the reaction time.

Backchannels that happen within this interval are almost all generated as a response to gaze shift. Backchannels that happen within a greater range may have occurred as a reaction to some other event. Since we can not be sure that no other factors influence the behavior of the listeners smaller intervals have been chosen to ensure ourselves no other factors play a critical role.

**Figure 3: The amount of backchannels that have occurred within a range of 100ms after some amount of time in milliseconds**

Backchannels that happen within this interval are almost all generated as a response to gaze shift. Backchannels that happen within a greater range may have occurred as a reaction to some other event. Since we can not be sure that no other factors influence the behavior of the listeners smaller intervals have been chosen to ensure ourselves no other factors play a critical role.

The graph also shows that after 500ms the chance a backchannel occurs is bigger than the average chance it occurs. Perhaps most listeners are not that fast in responding to a speaker.

### Remarkable is that these chances do not stay constant. They tend to increase during most of the conversations in the corpus. This accounts as well for the listeners as the eavesdroppers. In Table 2 the numbers of backchannels are noted. An increase of
the number of backchannels respectively 20% and 16% is measured if we compare the first with the second half. This is mainly because in the last 2% of the conversation suddenly a huge amount of backchannels occur. These backchannels were mostly not initiated by gaze but perhaps by prosody features or just the fact that the speaker was about to end his story.

Table 2 also shows the fact that listeners do respond more than eavesdroppers. In total 1007 backchannels were made by listeners. The eavesdroppers made 1760 but since they are with twice as much we should divide by 2 which makes only 880. So a listener gives average 14% more backchannels than an eavesdropper does.

Table 2: Amount of backchannels

<table>
<thead>
<tr>
<th></th>
<th>Listeners</th>
<th>Eavesdroppers</th>
</tr>
</thead>
<tbody>
<tr>
<td>First half</td>
<td>454 (908)</td>
<td>815</td>
</tr>
<tr>
<td>Second half</td>
<td>553 (1106)</td>
<td>945</td>
</tr>
<tr>
<td>Difference</td>
<td>99 (198)</td>
<td>130</td>
</tr>
<tr>
<td>Increment</td>
<td>0.2181</td>
<td>0.1595</td>
</tr>
</tbody>
</table>

Table 3 shows the chance a backchannel occurs within 2000 milliseconds after gaze in the first and last half of a conversation. The value of 2000 milliseconds is chosen since this lies between 500 and 2500 ms, the interval in which that chance a backchannel occurs after a gaze shift is the largest. Another value that lies in this interval would also suffice. Again we see a difference between the both halves although this difference is smaller.

Table 3: Chances a backchannel occurs within 2 seconds after gaze shift

<table>
<thead>
<tr>
<th></th>
<th>Listeners</th>
<th>Eavesdroppers</th>
</tr>
</thead>
<tbody>
<tr>
<td>First half</td>
<td>0.2612</td>
<td>0.2085</td>
</tr>
<tr>
<td>Second half</td>
<td>0.3210</td>
<td>0.2505</td>
</tr>
<tr>
<td>Difference</td>
<td>0.0598</td>
<td>0.0421</td>
</tr>
<tr>
<td>Increment</td>
<td>0.2288</td>
<td>0.2018</td>
</tr>
</tbody>
</table>

For clarification Figure 4 shows the development of the amount of backchannel occurrences during a conversation. For this diagram all kinds of backchannels were used, backchannels that are triggered by gaze and backchannels which are not. These backchannels could be for example trigger by prosody features, linguistic features or perhaps gaze or some other features. For the moment this is not important to know.

As we see this change a backchannel continues occurs increases slightly during the time. In the very last moments of an average conversation this chance increases a lot. This make sense since at the end of a conversation multiple factors that influence the backchannel behavior are present such as that the end of a sentence is reached and even the end of the whole conversation.

The shortest distance between two backchannel starts made by one person is 178ms. Since we must not match the other backchannel 177ms is a save choice. The problem with this choice is that this short distance between backchannel appear to happen very rarely. The good thing about this choice is that it shows which backchannels overlap exactly. It is very certain these backchannels respond to the same speaker action. This only happens in 13% of all cases as we see in Table 4. This percentage is not much. If we enlarge this range the percentage increases and it seems like there is more overlap but now it is possible we count wrong backchannel moments.

Figure 4: Amount of backchannels during a conversation

The vertical axis does not show the chance a backchannel occurs but the number of backchannels per 2% of a conversation. Every conversation was split in 50 parts and the amount of backchannels was measured there.

Figure 5: Chance a backchannel occurs within 2 seconds after gaze shift

The dark line represents the listeners actions and the lighter line the eavesdroppers. Most of the time eavesdroppers react less likely. The difference increases during the conversation.

The same is noticed if we only take backchannels that occurred directly after a speaker started gazing. Figure 5 shows the development of the chance a backchannel occurs within 2 seconds after gaze shift.

The pattern of this graph has lots of similarities with the graph of Figure 4. The surprising ending is again the eye catcher. The slow start is also seen in the graph before this one. We also see the difference between the eavesdroppers and the listeners.

4.3 Backchannels overlapping

Since we have multiple backchannel reactions on the same speaker the question arises if these backchannel patterns look the same. Do the listeners generally give backchannel at the same moment?

One important question that arises is: what is the same moment? Exactly the same looks impossible. But within what range can we say it is at the same moment.

There is a difference between a backchannel that happens as a reaction of the same action of a speaker and a backchannel that happens at the same time. It is possible a response will take place somewhat later than another response but both reacting at the same event.

Table 4: Overlap in backchannels between eavesdroppers and listeners

<table>
<thead>
<tr>
<th></th>
<th>100ms</th>
<th>177ms</th>
<th>200ms</th>
<th>400ms</th>
<th>800ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eavesdroppers</td>
<td>0.08</td>
<td>0.13</td>
<td>0.15</td>
<td>0.28</td>
<td>0.44</td>
</tr>
<tr>
<td>Listeners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Another way to find overlapping backchannels is to count what percentage backchannels happen during the same gaze period and thus also after the same gaze shift. Before we can calculate this it is useful to know what percentage backchannels starts actually while the speaker is gazing. In 1992 of 2767 cases (72%) a backchannel started while the speaker was gazing. When we also take the backchannels that happened 100ms before or after gaze we measure 2212 backchannels that started while the speaker was gazing (80%). It seems very important for a listener to response while the speaker is watching.

With this other way to match overlap it is found that in almost 40% of the cases a listener and a eavesdropper respond during exactly the same gaze moment. These backchannels happen approximately 750ms after each other. This is a much higher percentage compared with the 13% overlap found before.

This method also has its limitations. Two backchannels that happen during the same gaze period could be triggered by the same gaze shift but it is also possible one is triggered by a gaze shift and another by some other event.

Both methods showed that not even half of the backchannels of the eavesdroppers overlap with the backchannels of the listeners. So most of the backchannels made by the eavesdroppers are not seen at all by the speaker. Only the backchannels that do overlap the listeners backchannels are 'seen' by the speaker and could be used by the speaker in the rest of his speech.

The meaning of a backchannel made by an eavesdropper might be slightly different than the one made by the real listener. The exact contents of a backchannel might be slightly different. A backchannel could be made with much confident or sometimes it could be very uncertain. It seems likely a speaker will react differently in these situations but this research has not focused on this.

What happens when a backchannel made by an eavesdropper is not seen? Backchannels are useful for the speaker so he could know if he is understood. If a backchannel does not arrive the speaker, he might think the listener has problems understanding something and he may reformulate his words or he keeps asking for a backchannel. This might confuse the eavesdropper and may make him uncertain. Further backchannels are less likely to happens since a backchannel does not have the expected effect.

### 5. Evaluation

In this paper we have written about the relation between gaze shifts and backchannel-continuer occurrences. It is shown that such a relation exists. If a speaker makes more gaze shifts per minute it is likely this will trigger the listener to give more backchannel-continuer. After a gaze shift an listeners tends to give more backchannels. Especially between 500ms and 2500ms right after gaze shift.

In the overall analysis listeners do give more backchannels then eavesdroppers do. Even in the first half of the conversation this is seen.

Furthermore, this paper has focused on the development of the chance a backchannel occurs. We found out that the chance a backchannel may occur does not stay constant. In the MultiLis Corpus we used, this chance increased during the conversation.

What is perhaps most surprising is that the increment rate of the chance differs between the listeners and the eavesdroppers. Both groups do backchannel more at the end of a conversation in compare to the beginning. In comparison with eavesdroppers listeners tend to increase this rate faster. This also accounts for the chance a gaze shift triggers a backchannel. This chance increases a little bit during the conversation and again this increment is relative and absolute bigger with the listeners.

The same is noticed with backchannels made by a listener as a reaction to gaze shifts. We see similarities between the development of the chance these specific kind of backchannels and the backchannels in general.

Finally, the overlap between backchannels made by the listeners and eavesdroppers is measured. The backchannels given by the eavesdropper do not overlap in most cases. For sure we can only say a minimum of 13% and a maximum of 40% of the backchannels made by the eavesdroppers are the same backchannels as the ones made by the listeners.

In this study no other factors then gaze shifts have been studied. Consequently some conclusions cannot be made.

As mentioned before a difference exist in the individual behavior of all participants. In the experiment used in the corpus all participants acted as a speaker and a listener one time and as an eavesdropper two times. This made the difference between the individuals less important.

### 6. Discussion

Backchannels mostly happen during a speaker gazing and the chance a backchannels happens increases immediately after a speaker makes a gaze shift. An explanation for this is that a listener waits for giving feedback to the listener. When this moment finally arrives he will not wait and immediately throws a backchannel to influence the speaker.

Why do the results of the eavesdroppers and the real listeners differ? The major difference between both is that listeners are seen by the speaker and eavesdroppers are not. It follows that the speaker does only have the opportunity to adapt to that listener. All backchannels made by the listener could be interpreted and the speaker could react to this. He can chose for example to reformulate his sentences or give some extra clarification to the listener.

Backchannels made by the eavesdroppers are not seen by the speaker and he thus acts like he ignores most of the backchannels. Only the backchannels made at the same time and following the same event could somewhat be interpreted if we assume the backchannel response contains the same information.

Due to this fact an eavesdropper might feel backchanneling does not have the expected effect and may doubt the next time he will backchannel. This causes an eavesdropper to backchannel less. During the conversation he starts to doubt more and more and gives even less backchannels comparing to a real listener.

It's like a pupil says 'I understand' to his teacher. Sometimes he hears his pupil and continues his story. But pretty often he does not react to his pupil and keeps telling things the pupil is not interested in anymore. When this happens a lot it discourages the pupil for saying 'I understand' in the future since it has not much effect apparently. The same accounts of the backchanneling process.

I must agree that this conclusion lacks the certainty I wish it had. Because of the relative small group of participants there is room for randomness in the data and further experiments are needed to prove this conclusion. Possibly a larger difference between listeners and eavesdroppers is seen in other types of conversations where more backchannels are asked.

The differences between listeners and eavesdroppers, probably does not only have to do with backchanneling. Since the first backchannels, and thus the speaker was not able to react on any backchannel, happens with a smaller change with eavesdroppers then with real listeners.
7. Conclusions
Through this study we sought to answer two questions. First, whether and how gaze shifts influence the backchannel behavior in the corpus used. Second, whether the speakers response to backchannels does influence the listeners backchannel behavior.

In this research we have seen gaze shifts influences the backchanneling behavior. After a gaze shift a backchannel is more likely to occur. There is a clear relation between the amount of gaze shifts and the amount of backchannel moments. The chance a backchannel occurs is related to the reaction of the speaker to the listeners. If a listener does react on users backchannels he can expect more backchannels from the listener. The eavesdroppers in the experiment started with backchanneling like the normal listeners but during the conversation the behavior of the eavesdroppers and the listeners started to differ more and more. The only conclusion that make sense is that it is important for a speaker to react on the listeners behavior. The backchannel behavior may be an important aspect of this. Listeners give more backchannels when they notice it is useful.

Backchannels must be seen. A listener tries to backchannel on moments when the speaker gazes. When this does not work in most case, as is the case for eavesdroppers, backchannels happen less in the future.

These results could be used to improve the backchannel prediction models. This research showed that not only direct signals are important for predicting upcoming backchannels, but also the handling of the backchannel feedback by the speakers.

8. REFERENCES