Analysis of Non-Verbal Behavioural Cues in Phone Calls

Fara Stringfellow

School of Computing Science

Sir Alwyn Williams Building

University of Glasgow

G12 8RZ

A dissertation presented in part fulfilment of the requirements of the Degree of Master of Science at the University of Glasgow

Submitted: 21st September 2020

Abstract

This report is an observational study on the frequency and duration of non-verbal behaviour. It studies back-channel, filler, laughter, overlap and silence. This study is focused on mobile phone conversations between 120 unacquainted participants. These non-verbal social signals are relevant to enriching the understanding of human behaviour and allow for further research into the automatic detection of such behaviour.

This study shows statistically significant differences in the distribution of the frequency and duration of these non-verbal social cues depending on the participants role(caller/receiver), gender and opposite participants gender.

Education Use Consent

I hereby give my permission for this project to be shown to other University of Glasgow students and to be distributed in an electronic form.

Name: Signature:

Acknowledgements

I would like to express my sincere gratitude to Professor Alessandro Vinciarelli for his support and guidance in the different stages of this project.

Contents

Chapter 1 Introduction

Communication has become increasingly held over a variety of mediums distancing the individuals participating. Mediums such as video calling, phone calling and direct text messaging, each of which incrementally exclude sensory information that can be beneficial in conveying social behaviour. These mediums are useful in allowing for more frequent communication among people at a distance and are now considered a standard form of modern communication.

In person communication allows for the greatest variety of cues from facial expressions to body language. Limiting social cues to audio only creates difficulty in appropriately expressing emotions. Vocalising information in speech along with social context restricts individuals and in some instances a decision must be made to prioritise speech or a display of non-verbal cue.

This study examines a corpus of 60 phone call conversations between two unacquainted participants. The analysis focuses on the distribution of the non-verbal cues of back-channel, filler, laugher, overlap and silence. These cues appear most often in conversation and are easily distinguishable with well-documented literature on the social meanings and intentions of each. Further understanding the inferences of non-verbal cues provides insight into the field of social signal processing (Vinciarelli et al., 2012). The potential for automatic classification of social signals also aids in the field of Human-media interaction (Nijholt, 2014). Human-media interaction research is concerned with understanding the user's human behaviour by means of decoding social cues and generating a suitable response from a system.

Chapter 2 Background

The use of technology in communication reduces the individual's ability to communicate with non-verbal cues. The dependence on mobile devices as a means of communication has increased in modern decades (Toda et al, 2006). This analysis focuses on phone conversions therefore there is no means for visual cues such as facial expressions. To supplement the meaning behind words being spoken non-verbal audible cues are used. People are also unaware of instinctive spontaneous cues that may occur which could show insight into particular emotions being expressed.

There is increasing demand for technology to be socially intelligent. The ability for technology to automate the categorisation of social behaviour requires further analysis of understanding the nuances of behavioural cues that humans can pick up on. If technology was capable of analysing psychological behaviour efficiently, it could increase the ability for people to gain evaluation of their behaviour (Pentland, 2005). The future of computers and the interfaces that occur between it and humans will demand more than linguistic analysis of the words communicated. Human behaviour and our understanding of it are bound by current research methods which are slow and are the potential source of lack of standardisation and error. Computers are socially ignorant and the future of computing demands more of their real-time understanding of humans than that is currently present.

2.1 Non-verbal Cues

Non-verbal behaviour is an example of behaviour where social signals may add additional information into the analysis of the behaviour of an individual. Technology may be able to understand the words expressed in human interaction, and increasingly better understanding of the precise meaning of the words communicated is being developed. What offers a great sense of insight to a person's behaviour however is often held in their use of social signals and in this analysis specifically focusing on non-verbal cues. (Buck, VanLear, 2002). Specifically, the use of non-verbal cues can indicate certain signals such as dominance or hesitation or emotions such as pleasure.

As a listener non-verbal behaviour often provides necessary feedback in signalling their participation and intentions in conversation. In regarding all types of non-verbal behaviour, there is an extensive variety in ways in which individuals can chose to express themselves. There are however differences in norms among groups in the behavioural choices (Stubbe, 1998). Non-verbal communication has been described as mostly intentional with collective consensus on the social interpretation (Burgoon, Buller, and Woodall, 1996).

If verbal and non-verbal behaviors align they enhance the effectiveness of conveying their intended meaning. If they do not align they may express less collaborative signals such as lying or unease. Non-verbal behaviour conveys supplementary meaning to the information being expressed by the sender. Non - verbal behaviours may take prominence in adequality communicating the message than the information spoken.

The specific attributes of social signals are often deemed insignificant as opposed to the meanings they contain. In decoding the specific cause and implications of social signals much more about behavior can be observed in a single interaction than an analysis of the content of the discourse (Owren and Bachorowski, 2003)

2.1.1 Back-Channel

The use of back-channel i.e. worlds such as “yeah” and “uh-huh” suggest the listener is communicating agreement, understanding or appreciation to the speaker. Back-channel communicates supportiveness towards the speaker and is a social signal that is used to express engagement from the listener (Forbes and Cordella 1999). The use of back channel is generally considered not to be an intrusive cue with Bilous suggesting that the listeners use of backchannel is not an attempt to hold the floor(1988).

2.1.2 Filler

The use of fillers i.e. words such as “ehm”, “ah-ah” indicate a lack of ability to immediately express what is intended and that consideration and planning for the intended communication is occurring. Fillers also may express a means to maintaining the floor (Wennerstrom and Siegel, 2003).

2.1.3 Laughter

Laugher is a vocalisation that occurs almost always in a social environment. People are 30 times more likely to laugh in a social setting than alone (when omitting external entertainment media as a source) (Provine & Fischer 1989) . Laughter is the most diverse social signal in this analysis in terms of what messages it can convey; for instance, amusement or embarrassment (Morreall, 1982) . Laughter can much like back-channel indicate a listener’s participation in the conversation whilst allowing the floor to be shared (Vettin and Todt, 2004).

Laughter has been studied where the participants are directly responding to jokes and comedy , this is only representative of a laughter in response directly to humour and differs from the natural laughter emitted in social settings (Provine, 1996). Provine examined how speech takes priority over laughter, specifically how laughter occurs in pauses or at the end of sentences (1993).

2.1.4 Silence

The analysis of silence in this study measures pauses when the same participant is holding the floor. This may indicate hesitation, planning and anxiety in the participants use of silence.

2.1.5 Overlap

Overlap is the analysis of both participants speaking at the same time. In this analysis the overlap is attributed to the interrupter.

In literature back-channel and overlap are often associated and similarly categorised in the same research and the terminology can be confusing. For this analysis back-channel is specifically the utterances of reassurance terms(“yeah” ect.) while overlap Is any instance when both participants talk or makes a vocal cue in the same time interval

Chapter 3 Research Questions

Does the participant's role of caller or receiver affect the frequency or duration of the non-verbal cues ?

This question is influenced by the future needs of creating a more effective virtual avatar interaction. The differences in the roles of participant in this case could represent nuances in the social balance between the conversation initiator and participant.

Does gender affect the frequency or duration of the non-verbal cues ?

This question examines the well-documented literature studying of male and female sociolinguistics. This is a popular research topic as the differences in social behaviour gender provide insight that is enduringly relevant.

Does the frequency or duration of non-verbal cues differ when participants are paired with the same sex or a different sex ?

Often the social behaviour research focus on gender makes inferences that are not general to males and females in all conversations but who they are talking to influences their behaviour depending on status, age etc. Using this dataset further questions could be asked specific to each cue which takes into account the gender of the person the participant is speaking with.

Chapter 4 Experiment

This analysis was conducted on the SSPNet Mobile Corpus (Vinciarelli et al. ,2015)

The corpus is a collection of 60 phone calls of 120 participants. The participants were instructed to participate in a Winter Survival Task. The data collected is a continuous timeline of each call where the occurrence of backchannel, filler, laughter, overlap and silence were recorded for the caller and receiver.

The winter survival task that the participants were involved in required the participants to negotiate an item from a list of 12 that would increase the chances of survival in the aftermath of a plane crash. The list of 12 survival items were : steel wool, axe, pistol, butter can, newspaper, lighter without fuel, clothing, canvas, air map, whisky, compass, chocolate.

4.1 Participants

There were 120 participants. Of these participants 57 were male and 63 were female. There were 14 phone calls with a male only participants , there were 17 phone calls with a female only participants. There were 29 phone calls with a male and female participant. The participants were of British nationality.

Chapter 5 Analysis Methodology

5.1 Statistical Analysis Methodology

In all comparisons two groups were compared. The group variables being Male-Female, Caller-Receiver, Female talking with same vs different sex and Male talking with same vs different sex. The research hypothesis in all analysis being that there is a difference in the two groups cue quantifier.

For example, in the analysis of non-verbal cue duration which anaslysed gender effects, the following null and research hypothesis were used:

> **Null Hypothesis :** Male and Female subjects display the same mean duration of back-channel, laughter, overlap, silence or filler.
>
> **Research Hypothesis :** Male and Female subjects do not display the same mean duration of back-channel, laughter, overlap, silence or filler.

Two statistical methods were used to analyse the distribution of the non-verbal cues. For the continuous variables of cue density and duration a Student T-test is used. For the discrete variable of cue frequency a Chi Square test was used.

5.1.1 Chi-Square Test

In measuring the differences observed in two sample populations is likely that one group will incur a higher frequency a non-verbal cue than the other by chance. The chi-square test analyses whether the deviation from what was observed and what was expected by chance are significant enough to conclude that the differences in sample population did not occur at random.

The methodology for the chi-square analysis occurred by separating the corpus in two groups dependent on the variables of gender and role. The groups were divided so as the population proportions summed to equal the whole corpus. A chi- square test was not conducted on the same gender vs different gender participant analysis in section 6.3.

$$X^2 = \sum \frac{(O - E)^2}{E}$$

Where O is the observed value of the cue frequency and E is the expected value. The Expected value was calculated depending on the variable. For Caller and receiver, the participants shared equal portions of the call therefore the total cue frequency was calculated with the talk proportion of n = 0.5. For the Male Female analysis, the expected values were calculated in relation to the talk distribution. Male participants spoke for 52.8 % of the call time and female participants spoke for 47.2 % of the time. The expected values were therefore calculated:

$$E = n \, . \, F$$

Where F is the total cue instances per call and n is the relevant variables expected proportion where :

Caller :	n = 0.50
Receiver:	n = 0.50
Male:	n = 0.528
Female	n = 0.472

The X^2 value is then calculated and the associated p-value corresponding to the X^2 distribution. The critical value (α) is calculated controlling false discovery rate using the Benjenmini-Hochber procedure explained in section 5.1.3 (McDonald, 2009.). This procedure gives a unique critical value to each nonverbal cue. The degrees of freedom for all analysis is 1 as there are always 2 variables being analysed. Degrees of freedom is calculated by k -1 where k is the number of variables. The associated p-value for 1 degree of freedom in the chi – square distribution is estimated. If the p-value is less than the critical value, then the null hypothesis is rejected and the results deemed statistically significant.

5.1.2 Student T-test

The student t-test is used to determine if the difference in the means of two sample populations are significant enough to not have occurred by chance. This test is conducted by the ratio between the means of the two groups and the variance of the two groups. i.e. this is a ratio of the difference between the two groups and the difference within the two groups.

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

$$\bar{X}_1 = group\ 1\ mean$$

$$\bar{X}_2 = group\ 2\ mean$$

$$s_1^2 = group\ 1\ varience$$

$$s_2^2 = group\ 2\ varience$$

$$n_1 = number\ of\ participants\ group\ 1$$
$$n_2 = number\ of\ participants\ group\ 2$$

The t value is then calculated and the associated p-value corresponding to the t distribution is found. The critical value (α) is calculated controlling the false discovery rate using the Benjenmini-Hochber procedure explained in section 5.1.3 (McDonald, 2009.)

Degrees of freedom is determined by:

$$n_1 + n_2 - 2 = 120 - 2 = 118\ df$$

As the male and female sample groups are different, this is an analysis of two independent groups (or unpaired analysis). The associated p-value for 118 degree of freedom in the t distribution is estimated. If the p-value is less than the critical value, then the null hypothesis is rejected and the results deemed statistically significant.

In the analysis in section 6.3 in the analysis of same vs different gendered participants, the populations do not represent the whole corpus. The degrees of freedom therefore altered to :

$$Male\ analysis: n_1 + n_2 - 2 = 57 - 2 = 55\ df$$

$$Female\ analsyis = n_1 + n_2 - 2 = 63 - 2 = 61\ df$$

5.1.3 Critical Value

For all analysis the critical value is chosen to be 0.05 in a two -tailed analysis (i.e. 0.025 on either side of the distribution) A two- tailed analysis was chosen as upon initial analysis it was not assumed for any group to be specifically greater in their quantifier than the other, rather a difference in the 2 groups is being analysed instead. In this context for example it was not assumed that females laugh more frequently than males, but rather that either group may have a significant deviation from the other. This ensures that the critical value is not unofficially 0.1 from selectively choosing what group to pick as the greater cue emitter based on the means after they have been observed.

In measuring each variable, the critical value was adjusted to control the false discovery rate using the Benjenmini-Hochber method. For each non-verbal cue the resulting T-test and X^2 distribution p-values were ranked from lowest to highest then using:

$$\alpha = \frac{r}{m} Q$$

Where r is the cues ranking position, m is the number of variables being tested, for all instances of non-verbal cue analysis there are 5 non-verbal cues so m=5. Q is the selected critical value being 0.025. This generated a unique critical value for each non-verbal cue.

5.2 Measuring Cues

The nonverbal cues were analysed by cue frequency, cue density and mean cue duration. The cue frequency measures the number of instances a participant emits the non-verbal cue during the call. The cue density measures the frequency of cues emitted during the call divided by the total call length (i.e. cue occurrences per minute). The mean cue duration is the sum of individual cue lengths divided by the total cue frequency for that participant.

The cue frequency is non-relative to the duration of the call however when directly comparing two groups this is a valid quantifier in a Chi-Squared analysis which analyses discrete variables. The added analysis of cue density however measures frequency relative to the duration of the call. Both density and frequency must be statistically significant to conclude that non-verbal cues frequency is overall determined statistically significant.

Measuring frequency and duration may offer different insights into the behaviours linked to the cues. For instance, a higher frequency of periods of silence when one person is speaking may be due to the persons unique speaking style yet long instances of silence may be due to a higher sense of confusion and more planning on the participants verbalisation.

5.2.1 Cue Data Recording

The non-verbal cues of fillers, backchannel and laugher were recorded in the analysis at every occurrence.

The occurrence of overlap was deemed to occur when both participants were either speaking at the same time or emitting a non-verbal cue during same time interval. The assignment of the participant where the overlap occurs was held to the 'interrupter' as in the participant who was not previously emitting a vocalisation. This method of categorisation is not an exact measure of deeming a participants action an interruption as very small time intervals may occur where the opposing participant emits a vocalisation when in fact the other participant may have been regarded as holding the floor; when an instance of overlap occurs just deeming the previous vocal emitter as the participant holding the floor is not an entirely accurate assumption. Further attempts to specifically categorise overlap and accurately identify the interrupter are however out of the scope of this project.

Silence in this analysis records the interval of same participant silence. Specifically, this measures the breaks when a participant maintains holding the floor in which silence occur.

Chapter 6 Results

6.1 Effects of Role

In each phone call there is a random role assignment of the Caller and Receiver. The caller makes the call and the receiver answers the call. The corpus here being divided in two by this distinction of Caller and Receiver. There is a 50% participant speech proportion expected by the two groups.

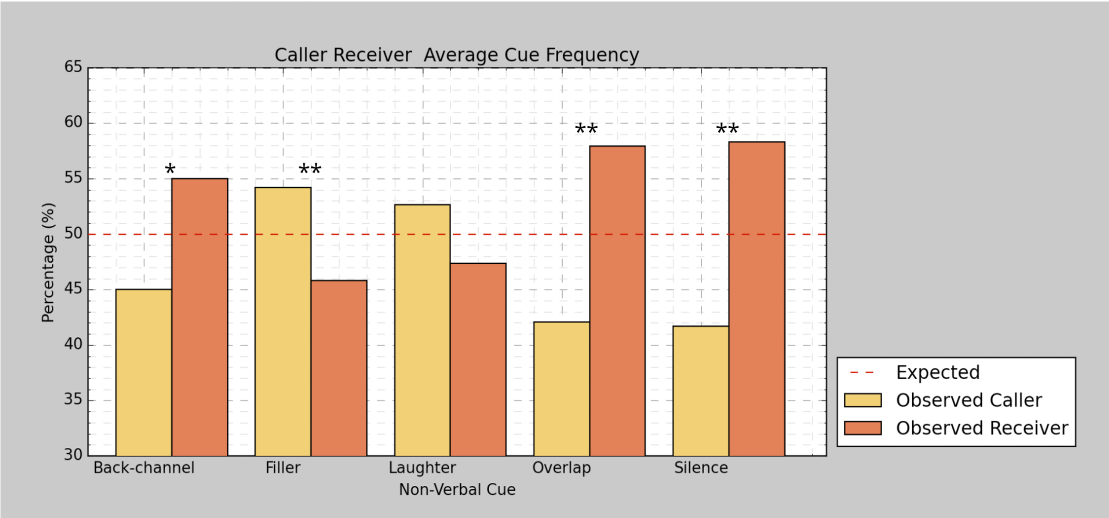

Figure 1: Chart shows the differences in distribution of cue occurrences depending on the role of the participants. Double asterisk indicates statistically significant deviation with a critical value of $\alpha = 0.01$, a single asterisk indicates $\alpha = 0.05$ (in a X^2 test with Benjamini-Hochberg correction).

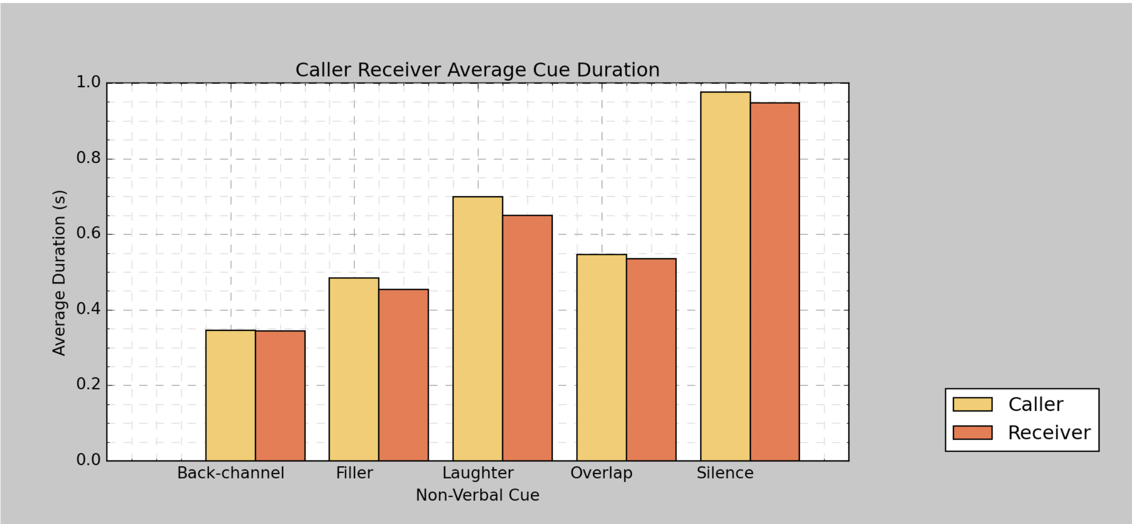

Figure 2: Chart shows the differences in distribution of cue duration depending on the role of the participants. No deviations were deemed statistically significant with a critical value of $\alpha = 0.05$ (in a two tailed independent Student T-test with Benjamini-Hochberg correction).

6.1.1 Back Channel

The back-channel density for Callers and receivers are relatively similar (0.063 and 0.716) respectively. The average duration is very similar with 0.699s for callers and 0.65s for receivers. Neither of these are statistically significant for the student t-test analysis. The Back-channel frequency analysis shows that callers emit 45.02 % of back-channel instances while receivers emit 54.98%. The frequency analysis is statistically significant according to the X^2 test however as the similar analysis of back-channel density is not statistically significant which is also a measure of the frequency of occurrence (but in relation to call length). The deviation in the caller and receiver back-channel frequency is not deemed a strong result.

6.1.2 Filler

The filler density for callers is 3.16 and 2.57 for receivers. The mean filler duration is very similar with 0.485s for callers and 0.454s for receivers. Neither of these are statistically significant in the student t-test. The filler frequency analysis shows that callers emit 54.2 % of filler occurrences while receivers emit 45.8 %. The frequency analysis is statistically significant according to the X^2 test however as the similar analysis of filler density is not considered statistically significant; therefore the deviation in the caller and receiver groups is not deemed a strong result.

6.1.3 Laughter

The laugh density for Callers is 1.36 laughs per min while 1.27 laughs per min for receivers. The average duration is 0.70s for callers and 0.65s for receivers. Neither of these are statistically significant in the student t-test. The laughter frequency analysis shows that callers produce 52.6% of occurrences and receivers produce 47.36% of occurrences. The laughter frequency analysis according to the X^2 test is also not statistically significant.

6.1.4 Overlap

The overlap density for callers is 4.054 overlaps per min while 5.404 overlaps per min for receivers. This is statistically significant in a student t-test analysis . The overlap frequency analysis shows that callers produce 42.1% of occurrences and receivers produce 57.9 % of occurrences. The overlap frequency analysis according to the X^2 test is statistically significant. The average duration is 0.55s for callers and 0.54s for receivers. While this duration analysis is not statistically significant in the student t-test what is notable is that there is a greater frequency of overlap occurrences in receivers to a statistically significant effect however the average duration of these overlap instances is less than that of callers.

6.1.5 Silence

The silence density for callers is 0.93 silence instances per min while 1.21 silences per min for receivers. This is statistically significant in a student t-test analysis . The silence frequency analysis shows that callers produce 41.7% of occurrences and receivers produce 58.3 % of occurrences. The silence frequency analysis according to the X^2 test is statistically significant. The average duration is 0.98s for callers and 0.95s for receivers. While this duration analysis is not statistically significant in the student t-test what is notable is that there is a greater frequency of silence occurrences in receivers to a statistically significant effect however the average duration of these silence instances is less than that of callers.

6.1.6 Role Effects Discussion

In the instance of back-channel it is not deemed in this analysis that there is an effect on the role of the participant and the frequency or duration of the occurrence back – channel. This suggests that in a scenario between a human speaker/initiator and an artificial responder/listener (e.g. a virtual avatar), the artificial responder should match back channel occurrences to that of the human when in discourse (Poppe, 2010).

The laugh density, frequency and duration analysis were all deemed to not affect the distribution of the caller and receiver groups.

Filler and same person silence suggest confusion and taking time to think and plan their speech. In the instance of filler, it is not deemed in this analysis that there is an effect on the role of the participant in the frequency or duration of the occurrence filler (as only the filler frequency was deemed statistically significant and not the density). This suggests that an artificial responder should implement a steady amount of filler alike to that of a human speaker.

It has been found that humans find words more identifiable if preceded by fillers such as 'uh' which validates that the inclusion of fillers should enhance the human users understanding of artificial speech(Tree, 2001). However Pfeifer and Bickmore found that there was no increase in satisfaction or naturalness when artificial responders implemented fillers (Pfeifer and Bickmore, 2009) .

Same-person silence is similar to filler and it can be associated with hesitations and the lack ability to immediately verbalise something rather than pausing to plan what to verbalise (Duhoe and Giddi, 2020). The effect of the roles of caller and receiver is deemed statistically significant on same-person silence. Receivers emit a higher frequency of silence whilst in their own stream of speech. This same-person silence suggests confusion and making time to think and plan speech which suggest that receiver feels more of a need to process their thoughts before expressing them. This difference may be amplified by the miscommunication issues associated with the audio only nature of a phone call as participants have less of an indication of when the speaker is finished.

What is interesting about the overlap distribution among the caller and receiver groups is that receivers emit a higher frequency of overlap with a statistically significant deviation, however the duration of these overlaps is deemed no different in the two groups. This amplifies the notion of subordination in the receiver and a need to be heard. However as these were voice calls there is the external factor of miscommunication due to a lack of clear audio output . This may reduce the participants understanding of when the other participant is speaking and when the speaker is finished. As the scenario in which this study occurs is solely audible output there is added confusion in that of the two participants and the receiver may deem that the caller has control and must interrupt more frequently to participate. Deeming overlaps as interruptions however is not necessarily true as Tannen suggests that overlaps allow for a more dynamic and animate conversation (Tannen, 1983).

For all 5 cues there were no deviations among the caller and receiver groups to suggest a difference in duration of the cue. This suggests the same intensity is observed in callers and receivers each cue occurrence. This may suggest that in the context of a virtual avatar that it should mirror the intensity of that of the human.

6.2 Effects of Gender

There were 57 male participants and 63 female participants of mixed pairing in call assignment. Male participants however spoke longer than females with calls with male participants maintaining 52.8 % talk time while females talked 47.2 % of the time.

The corpus here being divided in two by this distinction of Male and Female.

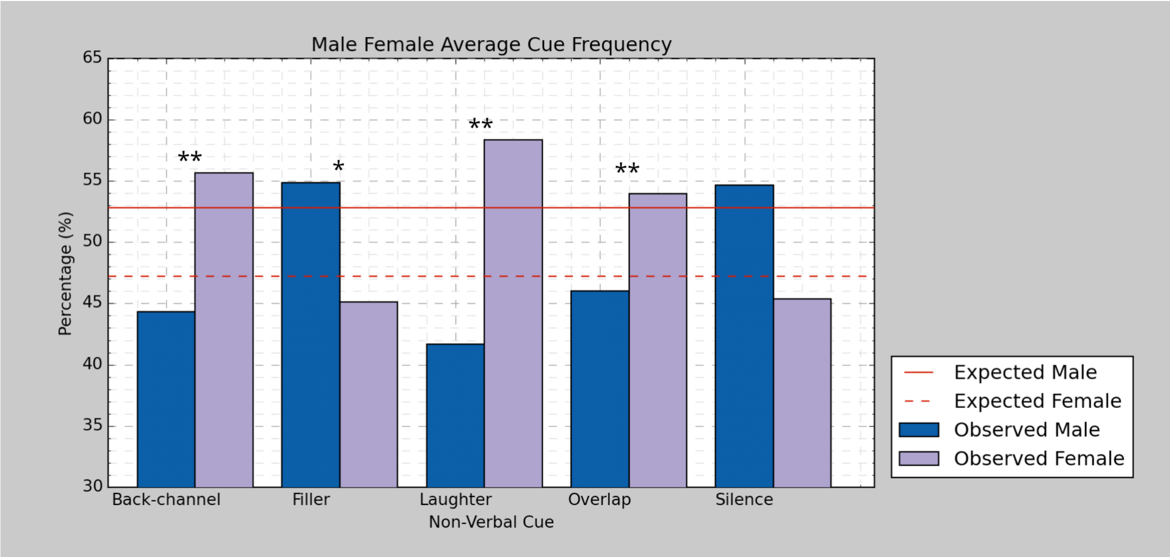

Figure 3: Chart shows the differences in distribution of cue occurrences depending on the gender of the participants. Double asterisk indicates statistically significant deviation with a critical value of $\alpha = 0.01$, a single asterisk indicates $\alpha = 0.05$ (in a X^2 test with Benjamini-Hochberg correction).

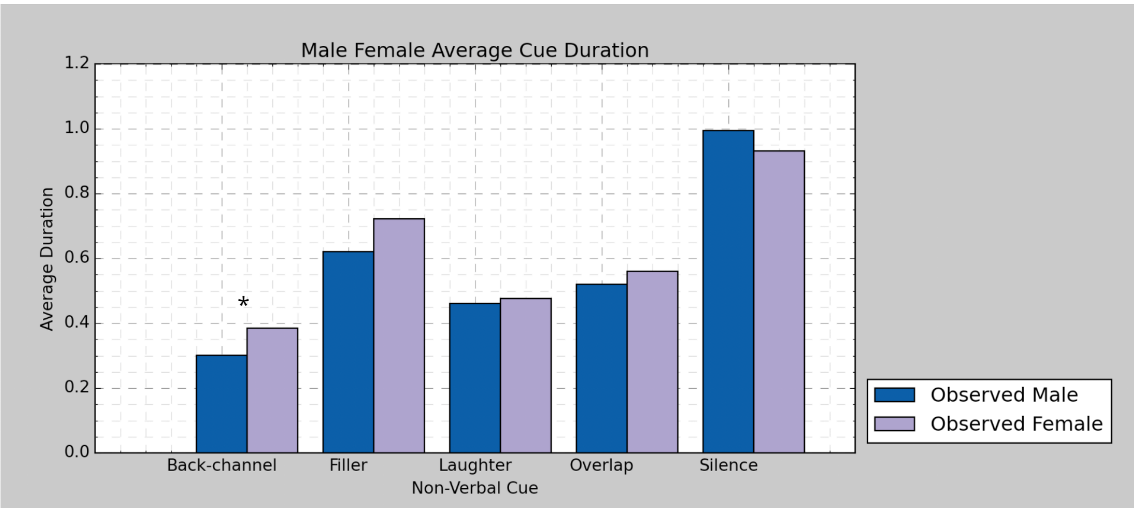

Figure 4: Chart shows the differences in distribution of cue duration depending on the gender of the participants. A single asterisk indicates statistically significant deviation with a critical value of $\alpha = 0.05$ (in a two tailed independent Student T-test with Benjamini-Hochberg correction).

6.2.1 Back Channel

The back-channel density for males is 0.524 instances per min while 0.808 instances per min for females. This is statistically significant in a student t-test analysis. The back-channel frequency analysis shows that males produce 44.3% of back-channel occurrences while females produce 55.67%. The back-channel frequency analysis according to the X^2 test is also statistically significant. The back-channel average duration for males is 0.30s for males and 0.38s for females. The difference between mean back-channel duration between males and females is statistically significant.

6.2.2 Filler

The filler density for males is 2.997 occurrences per min and for females is 2.750 occurrences per min for females. Males emit 52.8 % of filler occurrences while females emit 47.2 % of filler occurrences. The filler frequency analysis is statistically significant according to the X^2 test however the filler density analysis is not statistically significant according to the student t-test; therefore this is not a strong result. The mean filler duration for males is 0.46s and 0.48s for females, this is not statistically significant according to the student t-test.

6.2.3 Laughter

The laughter density for males is 1.003 laughs per min and 1.605 laughs per min for females. This is statistically significant in a student t-test analysis. The laughter frequency of shows that males produce 41.7% of laugh occurrences while females produce 58.3 % of instances. The male female laughter frequency analysis is statistically significant according to the X^2 test. The mean laughter duration for males is 0.62s for male and 0.72 for females. This is not statistically significant according to the student t-test.

6.2.4 Overlap

The overlap density for males is 4.091 occurrences per min and 5.306 occurrences per min for females. The overlap frequency shows that males produce 52.8 % of overlap instances and female produce 47.2 % of occurrences. Both overlap density and frequency are statistically significant according to the student t-test and X^2 test respectively .The mean overlap duration is 0.52s for male and 0.56s for females, this is not statistically significant according to the student t-test.

6.2.5 Silence

The silence density for males is 1.123 silence instances per min and 1.023 silence instances per min for females. The silence density is not statistically significant according to the student t-test. The silence frequency analysis shows that male produce 52.8 % of silence instances while females produce 47.8% of silence instances. The silence frequency analysis is not statistically significant according to the X^2 test. The mean duration for male silence is 1.00 s for males and 0.93s for females, this is not statistically significant according to the student t-test.

6.2.6 Gender Effects Discussion

This analysis does not suggest that the use of filler is dependent on gender. This is supported by Laserna et al. who found that filler words that facilitated pauses ("uhm", "uh") were not associated with gender (2014). The study did find that filler words that facilitated discourse ("like", "I mean") were associated with gender. They found that females used more discourse fillers. The distinction between discourse and paused fillers is out of the scope of this study. The use of discourse fillers indicates conscientiousness while paused fillers indicate reflection of anxiety (Christenfeld, Creager, 1996). The notable factors relevant to this study is that the participants were strangers and were engaged in a decision making and persuasion task. Fillers can be used as an intentional signal to inform the listener that they want to maintain the floor while they are deciding on words to say or fillers may be used to surrender the floor (Clark and Tree, 2002). Same-person silence is a similar social signal to filler in that it takes time for hesitation in speech, the effect of gender was not deemed statistically significant on same-person silence. The silence however in this analysis may be affected by the miscommunication associated with unclear turns in a phone conversation.

Females emitted a higher frequency and density of backchannel than males. This may be due to the conversational differences in males and females that indicate females are more friendly, collaborative and supportive (Zhang, 2014). Laughter similarly is a social signal that at times conveys the message of support to the speaker acting similarly to back channel and so the gender differences in back channel and laugher may be linked. Females also produced back-channel utterances for a longer average duration than men with a statistically significant deviation. This may be due to the greater degree of emotion and attentiveness that women express in conversation (Tannen, 1990).

The overlap distribution shows a statistically significant deviation with females showing a higher frequency of overlap than males. This conflicts with the meta-analysis conducted by Anderson and Leaper which researched gender differences in interruptions (1998). They found that overall men are more likely to interrupt than women. Although the categorisation of overlap does not explicitly represent an interruption its behaviour is linked. The only added nuance to this study specifically is that the conversation is occurring over phone calls and may incur a greater variation in overlap due to the miscommunication of turn-taking. Interruptions can be considered a sign of dominance or indifference (Occult and Harvey, 1985). The higher frequency in overlap may indicate that females may wish to ensure that they are heard; this is further backed up by the greater tendency for males to talk for a longer duration found in this study in section 6.4 . Overlap was considered any instance when two participants are vocalising at the same time interval so includes shared laugher and some instances of backchannel which are not interruptions in which the floor is competed for (Eecke and Fernandez, 2016). This analysis therefore does not definitively express the notion that females interrupt more than males. Refining the definition of overlap to only include speech and fillers would be a better representation for this particular investigation.

The laughter distribution among females and males is statistically significant. Females laugh more frequently than males and this corresponds to a well-documented observation in the literature (Provine, 1993; Owren and Bachorowski, 2003). The discussion of gender differences and laughter is covered in more detail in Section 6.3. Jefferson et al. argue that a laughs per minute quantification is not relevant unless supplemented with the context in which it occurs(1987). The monolithic categorisation of laughter does not offer the contextual differences of the situation such as how Laughter can express a funny remark and indicate sociable qualities contrasting to how laughter in other contexts may be construed as offensive.

6.3 Effects of Same Gender or Mixed Gender Conversations

This section analyses the cue distribution of participants depending on if they were in conversation with the same or different gender. The analysis counts cues of the individual participants depending on who they are speaking to rather than the cue distribution of both participants combined over the duration of the conversation.

For this analysis there were 17 female only calls, 14 male only calls, and 29 mixed sex calls. This divided the corpus into calls that contained at least one participant in the analysis and omitted calls with only opposite sex pairings.

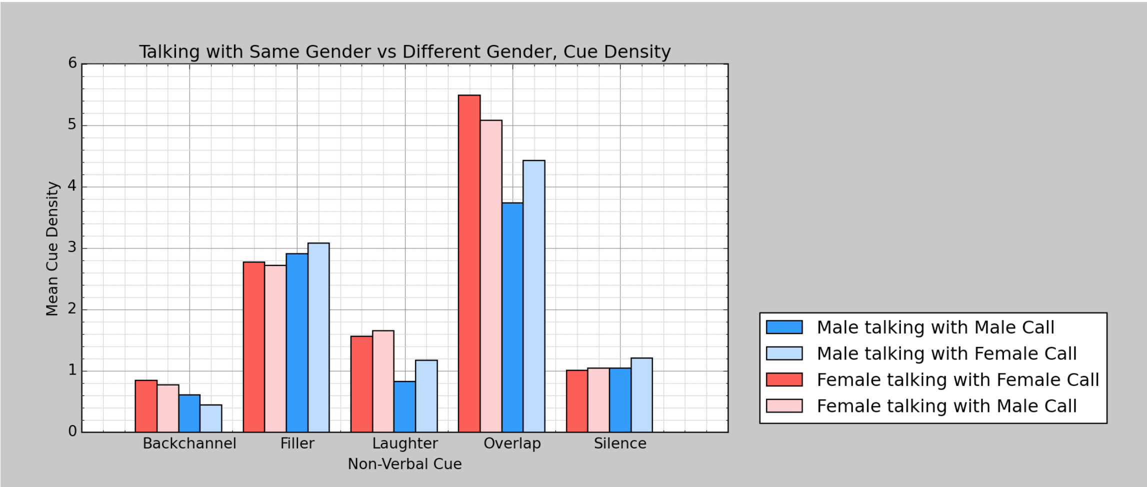

Figure 5: Chart shows the differences in distribution of cue density depending on the gender of the participant and the gender of who they are in conversation with. No deviations were deemed statistically significant with a critical value of $\alpha = 0.05$ (in a two tailed independent Student T-test with Benjamini-Hochberg correction).

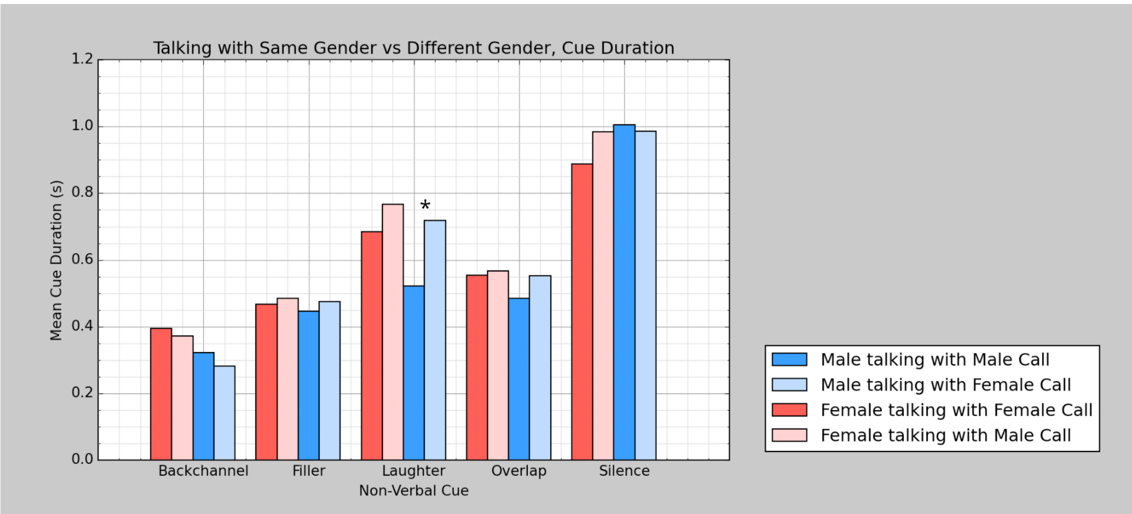

Figure 6: Chart shows the differences in the distribution of cue duration depending on the gender of the participant and the gender of who they are in conversation with. A single asterisk indicates statistically significant deviation with a critical value of $\alpha = 0.05$ (in a two tailed independent Student T-test with Benjamini-Hochberg correction).

It would be expected that there would be a higher frequency of female laughter when in conversation with a male than with a female. This would support the frequently observed results that male speakers are more likely to evoke a laugh from listeners than females (Provine, 1993). This analysis does not categorise specifically what the source of the laughter was and so such evaluations of the results are not comparable to works focused entirely on gender differences in laughter.

To support the common assumptions that men are better at getting laughs and females are more likely to laugh; however the results of this study appear to conflict with this notion. This may be due to the social situation of this study; the phone call setting is a less sensory social setting than in person. Most studies where this gender and laughter observation is made is conducted with people face to face. Owren and Bachorwski found that females laughed more frequently when in conversation with unacquainted males than with females however this experiment like most was in person (2003).

It has been suggested that this laughter balance is to promote their attractiveness towards males (Jefferson et al 1987). The lack of the visual sense of the other participant may have resulted in the female participants less inclined to communicate the meanings involved in the laugher to their male listeners or this may be due to men not feeling the need to encourage laughter in females as the interaction is so impersonal and their influence on the female participants might be less effective. It may be harder to decide if they are interested in a person over the phone , participants are more likely to leave the interaction feeling indifferent to one another than if they met in person.

Coates examined the connection between laughter and intimacy, in that a conversation can take the course of a serious or playful nature and this is dependent on the level of collaboration and solidarity among the speakers (2007). In conversation over phone calls there is a certain distance that limits the level of solidary and therefore intimacy among participants. This may indicate the lack of a statistically significant deviation in calls where a female is speaking to a male rather than a female.

What is additionally interesting about this observation is that even though not statistically significant females are more likely to laugh at females than men are to laugh at either females or males. It should be noted that the research conducted on gender dynamic of laugher reaches it conclusions based on a consensus that participants are heterosexual.

The laughter distribution shows a statistically significant deviation with males showing a longer duration of laughter when talking with females than males. This indicates a greater intensity of laughter when males are speaking with females which is closer to the average duration of all female laughter observed. This may indicate that they are matching the intensity to that of the female when talking with them.

6.4 Gender Effects on Call Duration

This section analysis the gender effects on the call duration and therefore the talk durations. The Corpus was split up the same as in section 6.3. There was no controlling of the false discovery rate as only one group comparison was made between male only calls and female only calls.

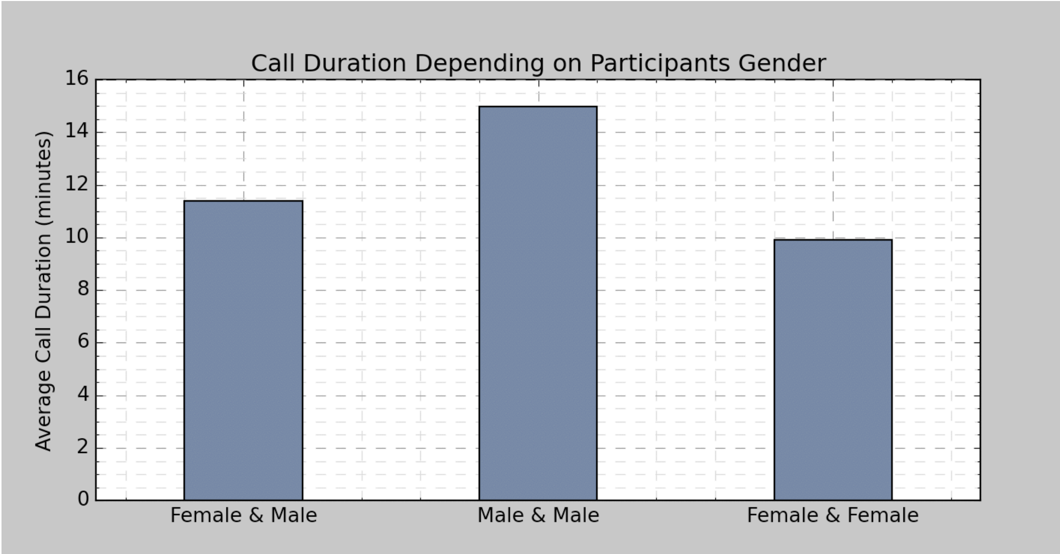

Figure 6: Chart shows the differences in the mean call duration depending on if the calls were mixed sex or same sex. It was found that the distribution of male only conversations and female only conversations was a statistically significant deviation with a critical value of $\alpha = 0.05$ (in a two tailed independent Student T-test).

The average call duration for Male: Male calls is 15.0 seconds while for Female: Female calls the average is 9.91 seconds. This is statistically significant using a two-tailed independent Student T-Test with a critical value of 0.05.

The difference in the male only and female only conversation times appears to object to the stereotype that women are more talkative than men, especially when talking with other women. There have been numerous studies which also indicate that men talk more than women (Spendor, 1980).

This may indicate that men are more likely to hold the floor than females. Holding the floor for longer has been linked to dominant personalities (Rogers and Jones, 1975).

Chapter 7 Conclusion

This study has analysed the distribution of non-verbal cues (back-channel, filler, laugher, overlap, silence) in the SSPNet Corpus of 60 phone calls between unacquainted participants. This analysis documents how the frequency and duration of these non-verbal cues differ depending on the role and gender of the participant and the gender of the other conversation participant.

Back-channel was shown to be dependent on role and gender with a higher frequency displayed by receivers and females. Additionally, females displayed a longer duration of back channel. Females have more supportive and attentive tendencies so a higher proportion of back channel than that of males is consistent with the literature. The greater frequency in backchannel in receivers was considered to be due to the complexities generated in the roles assignment with the caller deemed the initiator and therefore the conversation flow may put the receiver in a position of responding to the callers statements more frequently than the caller responding to the receiver.

Filler was shown to be dependant on role and gender with callers and males displaying a higher frequency of filler. Filler indicates hesitation in planning speech indicating in this analysis either anxiety or potentially a greater sense responsibility in completing the task appropriately.

Laugher was shown to be dependent on gender with a higher frequency displayed by females. Additionally, males displayed a longer duration of laugher when in conversation with females than with males. Laugher is a diverse social signal with many complex messages; however research commonly documents how females laugh more often than males which was also observed in this analysis. This analysis did not observe however that females laugh more frequently with males than with females which is a common observation in the literature.

Overlap was shown to be dependent on role and gender with a higher frequency displayed by receivers and females. Overlap is considered a measure for a participants tendency to interrupt the speaker which can be associated with dominant personalities. This common assumption conflicts with the results of this study perhaps due to the conversation differences in a phone conversation with receivers and females likened to less dominant characteristics and a need to be heard and compete for the floor.

Silence instances between same participant speech was shown to be dependent on the role of participants with receivers displaying a higher frequency of same person silence. This indicates a greater sense of hesitation and uncertainty in planning speech. This result was potentially due to the role of the receiver as most likely a responder to the caller's statements which demands a more immediate degree of improvisation.

Apart from male laughter duration, no non-verbal cues were shown to be dependent on whether participants where in conversation with the same or opposite sex. This was a surprising result with studies often noting the differences in social behaviour depending on the participants gender.

This study shows the clear importance of non-verbal social signals in phone conversations considering the lack of visual sensory feedback.

7.1 Future Considerations

This analysis is limited to vocal non-verbal behaviour which is somewhat distinctive in that it analyses phone call conversation in contrast to the majority of research which is focused on in person conversation and the effects on social signals. This measures specifically when interactions are limited to voice only and suggest insights into how groups use behaviour to express intentions. There are also various other non-verbal signals like body language and facial expressions that are not discussed in this report. A study directly comparing the differences between audio only and in person non-verbal signals could suggest the added significance put on to certain social signals in phone calls. Future work could also focus on if social behaviour is altered depending on the added visual communication of a video call.

Fruehwald and Laserna et al conducted studies and found that the use of fillers was dependant on age, this could be interesting to analysis the use of filler with this variable and a more diverse age range of participants (2016). Communication between humans is more frequently dependent on technology and interfacing with technology is a more regular occurrence for all ages. A study from an elderly sample demographic and their distribution on which social signals they prioritise will be influential in promoting inclusive design practices.

Enhancing digital simulation experiences or the interaction with a virtual avatar (or embodied conversational agent) is notes as a potential use for the caller receiver analysis section. A more focused study could be implemented to explicitly concentrate on assigning participants with roles of initiator and responder to fulfil this topic more directly .

Cultural differences were not explored in this study as all participants were of British nationality however research is more recently concerned with the differences in social signal distribution and their meanings.

Analysis of overlap on this study does not distinguish between overlap and interruptions. Categorising overlaps as interruptions by only including filler and speech can measure if there is any differences between the interrupter and the interrupted.

Chapter 8 References

Anderson, K.J. and Leaper, C., 1998. Meta-analyses of gender effects on conversational interruption: Who, what, when, where, and how. Sex roles, 39(3-4), pp.225-252.

Bilous, Frances R. (1988) "Dominance and Accommodation in the Conversational Behaviors of Same and Mixed- Gender Dyads" in, Language and Communication, vol. 8, 3-4, pp.183-194.

Buck, R. and VanLear, C.A., 2002. Verbal and nonverbal communication: Distinguishing symbolic, spontaneous, and pseudo-spontaneous nonverbal behavior. Journal of communication, 52(3), pp.522-541.

Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1996). *Nonverbal communication: The unspoken dialog*. New York: Harper & Row.

Christenfeld, N. and Creager, B., 1996. Anxiety, alcohol, aphasia, and ums. *Journal of personality and social psychology*, *70*(3), p.451.

Clark, H.H. and Tree, J.E.F., 2002. Using uh and um in spontaneous speaking. *Cognition*, *84*(1), pp.73-111.

Coates, J., 2007. Talk in a play frame: More on laughter and intimacy. Journal of Pragmatics, 39(1), pp.29-49.

Duhoe, A.A.A. and Giddi, E., 2020. Semantic Analysis of Silence in Conversational Discourse. *Journal of Linguistics and Foreign Languages*, *1*(1), pp.18-32

Eecke, P. and Fernández, R., 2016. On the Influence of Gender on Interruptions in Multiparty Dialogue. In Interspeech (pp. 2070-2074)

Forbes, Katie; Cordella, Marisa (1999) "The Role of Gender in Chilean Argumentative Discourse", in IRAL, vol. 37, 4, Nov. pp.277-289.

Fruehwald, J., 2016. Filled pause choice as a sociolinguistic variable. University of Pennsylvania Working Papers in Linguistics, 22(2), p.6.

Jefferson, G., Sacks, H. and Schegloff, E.A., 1987. Notes on laughter in the pursuit of intimacy.

Laserna, C.M., Seih, Y.T. and Pennebaker, J.W., 2014. Um... who like says you know: Filler word use as a function of age, gender, and personality. *Journal of Language and Social Psychology*, *33*(3), pp.328-338.

Morreall, J., 1982. A new theory of laughter. Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 42(2), pp.243-254.

McDonald, J.H., 2009. Handbook of biological statistics (Vol. 2, pp. 6-59). Baltimore, MD: sparky house publishing.

Mott, H. and Petrie, H., 1995. Workplace interactions: women's linguistic behavior. *Journal of Language and Social Psychology*, *14*(3), pp.324-336.

Nijholt, A., 2014. Breaking fresh ground in human–media interaction research. Frontiers in ICT, 1, p.4.

Orcutt, J.D. and Harvey, L.K., 1985. Deviance, rule‐breaking and male dominance in conversation. Symbolic Interaction, 8(1), pp.15-32.

Owren, M.J. and Bachorowski, J.A., 2003. Reconsidering the evolution of nonlinguistic communication: The case of laughter. Journal of Nonverbal Behavior, 27(3), pp.183-200.

Pentland, A., 2005. Socially aware, computation and communication. Computer, 38(3), pp.33-40.

Pfeifer, L.M. and Bickmore, T., 2009, September. Should agents speak like, um, humans? The use of conversational fillers by virtual agents. In *International Workshop on Intelligent Virtual Agents* (pp. 460-466). Springer, Berlin, Heidelberg.

Poppe, R., Truong, K.P., Reidsma, D. and Heylen, D., 2010, September. Backchannel strategies for artificial listeners. In *International Conference on Intelligent Virtual Agents* (pp. 146-158). Springer, Berlin, Heidelberg.

Provine, R.R. & Fischer, K.R. 1989;, "Laughing, Smiling, and Talking: Relation to Sleeping and Social Context in Humans", Ethology, vol. 83, no. 4, pp. 295-305.

Provine, R.R., 1993. Laughter punctuates speech: Linguistic, social and gender contexts of laughter. Ethology, 95(4), pp.291-298.

Provine, R.R., 1996. Laughter. American scientist, 84(1), pp.38-45

Rogers, W.T. and Jones, S.S., 1975. Effects of Dominance Tendencies on Floor Holding and Interruption Behavior in Dyadic Interaction1. Human Communication Research, 1(2), pp.113-122.

Spender, D. (1980). Man Made Language. London: Routledge.

Stubbe, M., 1998. Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. *Journal of Pragmatics*, *29*(3), pp.257-289.

Tannen, D., 1983. 10. when is an overlap not an interruption? one component of conversational style,". *Selected Papers*, *119*.

Tannen, D., 1990. You just don't understand: Women and men in conversation (p. 42). New York: Morrow.

Toda, M., Monden, K., Kubo, K. and Morimoto, K., 2006. Mobile phone dependence and health-related lifestyle of university students. Social Behavior and Personality: an international journal, 34(10), pp.1277-1284.

Tree, J.E.F., 2001. Listeners' uses of um and uh in speech comprehension. *Memory & cognition*, *29*(2), pp.320-326.

Vettin, J. and Todt, D., 2004. Laughter in conversation: Features of occurrence and acoustic structure. Journal of Nonverbal Behavior, 28(2), pp.93-115.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F. and Schroeder, M., 2011. Bridging the gap between social animal and unsocial machine: A survey of social signal processing. IEEE Transactions on Affective Computing, 3(1), pp.69-87.

Vinciarelli, A., Chatziioannou, P. and Esposito, A., 2015. When the words are not everything: the use of laughter, fillers, back-channel, silence, and overlapping speech in phone calls. Frontiers in ICT, 2, p.4.

Wennerstrom, A. and Siegel, A.F., 2003. Keeping the floor in multiparty conversations: Intonation, syntax, and pause. Discourse Processes, 36(2), pp.77-107.

Zhang, C., 2014. A Sociolinguistic Study of Miscommunication in Male-female Conversation. Asian Social Science, 10(17), p.240.

Appendix A Analysis Data of Cue Distribution

Caller vs Receiver

Caller Receiver Density			
Cue	**Caller**	**Receiver**	**T-test**
Back Channel			
Mean	0.6299635687222084	0.7162210620939732	0.7175
Variance	0.377569782791268	0.48953787224484774	
Laugher			
Mean	1.36877004	1.26858049	0.4904
Variance	1.34705096	1.15658648	
Filler			
Mean	3.1629503717025123	2.572356728712718	1.8793
Variance	3.172120184131926	2.753525044618561	
Overlap			
Mean	4.0543150464693705	5.403833009799322	3.535
Variance	3.428281297611415	5.316554849468991	
Silence			
Mean	0.9268119984616836	1.2142125241458417	2.086
Variance	0.5377108573592344	0.5973009817457632	

False Discovery Rate: Caller Receiver Density				
Cue in order of Rank	**T Test Value**	**T test p value**	**Rank**	**(rank/m)Q**
Overlap	3.535	0.0003	1	0.005
Silence	2.086	0.0196	2	0.001
Filler	1.8793	0.03133	3	0.015
Laughter	0.7175	0.2372	4	0.02
Back channel	0.4904	0.3159	5	0.025

Caller Receiver Duration			
Cue	**Caller**	**Receiver**	**T-test**
Back Channel			
Mean	0.3462	0.3442	0.0667
Variance	0.0305	0.0239	
Laugher			
Mean	0.6994	0.65	1.0784
Variance	0.0474	0.0783	
Filler			
Mean	0.4845	0.4538	1.9021
Variance	0.0085	0.0071	
Overlap			
Mean	0.5456	0.5357	0.4908
Variance	0.0121	0.0122	
Silence			
Mean	0.9768	0.9476	0.3928
Variance	0.1699	0.1621	

False Discovery Rate: Caller Receiver Duration				
Cue in order of Rank	**T Test value**	**T test p value**	**Rank**	**(rank/m)Q**
Filler	1.9021	0.0298	1	0.005
Laugh	1.0784	0.1415	2	0.001
Overlap	0.4908	0.3122	3	0.015
Silence	0.3928	0.3475	4	0.02

Back channel	0.0667	0.4735	5	0.025

Caller Receiver Frequency			
	Caller	**Receiver**	**Chi Square**
Back Channel			
Observed	457.0	558.0	10.0502
Expected	507.5	507.5	
Laugher			
Observed	950.0	855.0	5
Expected	902.5	902.5	
Filler			
Observed	2120.0	1792.0	27.501
Expected	1956.0	1956.0	
Overlap			
Observed	2738.0	3771.0	163.9405
Expected	3254.5	3254.5	
Silence			
Observed	580.0	811.0	38.3616
Expected	695.5	695.5	

False Discovery Rate: Caller Receiver Frequency				
Cue in order of Rank	**Chi square value**	**Chi p value**	**Rank**	**(rank/m)Q**
Overlap	163.9405	<0.00001	1	0.005
silence	38.3616	<0.00001	2	0.001
filler	27.501	<0.00001	3	0.015
Back channel	10.0502	0.001523	4	0.02
laughter	5	0.0253	5	0.025

Male vs Female

Male Female Density			
	Male	**Female**	**T-test**
Back Channel			
Mean	0.5238	0.8082	−2.4572
Variance	0.2667	0.5488	
Laugher			
Mean	1.0026	1.6047	−3.1316
Variance	0.5762	1.6919	
Filler			
Mean	2.9974	2.7502	0.778
Variance	2.8876	3.17	
Overlap			
Mean	4.0911	5.3063	−3.186
Variance	3.3387	5.4746	
Silence			
Mean	1.1228	1.0232	0.715
Variance	0.5399	0.6276	

False Discovery Rate: Male Female Density				
Cue in order of Rank	**T -test value**	**T test p value**	**Rank**	**(rank/m)Q**
Overlap	−3.186	0.0009	1	0.005
laughter	−3.1316	0.0011	2	0.001
BC	−2.4572	0.0077	3	0.015
filler	0.778	0.2191	4	0.02
silence	0.715	0.2380	5	0.025

Male Female Duration			
	Male	**Female**	**T-test**
Back Channel			
Mean	0.3017	0.3846	−2.8159
Variance	0.0301	0.0212	
Laugher			
Mean	0.6217	0.7227	−2.2028
Variance	0.082	0.0418	
Filler			
Mean	0.4615	0.4761	−0.8981
Variance	0.0065	0.0094	
Overlap			
Mean	0.5196	0.5596	−1.9965
Variance	0.0148	0.009	
Silence			
Mean	0.9953	0.9322	0.8579
Variance	0.1301	0.1968	

False Discovery Rate: Male Female Duration				
Cue	**T -test value**	**t-test value**	**Rank**	**(rank/m)Q**
Backchannel	−2.8159	0.00285	1	0.005
Laughter	−2.2028	0.01477	2	0.001
Overlap	−1.9965	0.0308	3	0.015
Filler	−0.8981	0.1855	4	0.02
Silence	0.8579	0.1963	5	0.025

Chapter 9

Male Female Frequency			
	male	**female**	**chi**
Back Channel			
observed	450.0	565.0	29.1841
expected	535.9200000000001	479.08	
Laugher			
observed	752.0	1053.0	89.8487
expected	953.0400000000001	851.9599999999999	
Filler			
observed	2146.0	1766.0	6.6409
expected	2065.536	1846.464	
Overlap			
observed	2996.0	3513.0	119.7563
expected	3436.752	3072.248	
Silence			
observed	760.0	631.0	1.8834
expected	734.448	656.552	

False Discovery Rate: Male Female Frequency				
Cue in order of Rank	**Chi square value**	**Chi p value**	**Rank**	**(rank/m)Q**
Overlap	119.7563	<0.00001	1	0.005
laughter	89.8487	<0.00001	2	0.001
Backchannel	29.1841	<0.00001	3	0.015
filler	6.6409	0.0100	4	0.02
silence	1.8834	0.1699	5	0.025

Call Duration

Call Duration Analysis			
	Mean duration (mins)		
Male & Female	11.38		
Male & male	15.0		
Female & Female	9.91		
	T-test value	T Test p value	Critical Value
Male: Male vs Female: Female Call Duration	2.077*	.023731	0.25 **
* 29 Degrees of freedom			
**Two-tailed t-test with $\alpha = 0.05$			

Same Gender vs Different Gender Conversation

False Discovery Rate: Male: Male vs Male: Female Density				
Cue in order of rank	**Student t-test value**	**t-test p value ***	**Rank**	**(rank/m)Q**
Laughter	1.7356	0.044	1	0.005
Overlap	1.4446	0.077	2	0.001
backchannel	1.1415	0.129	3	0.015
silence	0.8351	0.204	4	0.02
filler	0.4002	0.345	5	0.025
*55 degrees of freedom				

False Discovery Rate: Female: Female vs Male: Female Density				
Cue in order of rank	**Student t-test value**	**t-test p value ***	**Rank**	**(rank/m)Q**
overlap	0.6852	0.248	1	0.005
backchannel	0.3966	0.347	2	0.001
laughter	0.2599	0.398	3	0.015
silence	0.1991	0.421	4	0.02
filler	0.1272	0.450	5	0.025
*61 degrees of freedom				

False Discovery Rate: Male: Male vs Male: Female Duration				
Cue in order of rank	**Student t-test value**	**t-test p value ***	**Rank**	**(rank/m)Q**
Laughter	2.7477	0.0041	1	0.005
Overlap	2.1780	0.0169	2	0.001
filler	1.3904	0.0850	3	0.015
backchannel	0.8861	0.1913	4	0.02
Silence	0.1847	0.4270	5	0.025
*55 degrees of freedom				

False Discovery Rate: Female: Female vs Male: Female Duration				
Cue in order of rank	**Student t-test value**	**t-test p value ***	**Rank**	**(rank/m)Q**
overlap	1.5266	0.0668	1	0.005
backchannel	0.8460	0.2004	2	0.001
laughter	0.6786	0.2500	3	0.015
silence	0.6418	0.2622	4	0.02
filler	0.5104	0.3053	5	0.025
*61 degrees of freedom				