The Benefits of Context in Human Communication to Natural Language Processing

1st ___________ 2nd ___________ 3rd ___________

***Abstract*—Contextual information often aids the understanding of human communication, yet despite this, few speech recognition tools take advantage of context. While the benefits of context in human communication are rather obvious, the same cannot be said for human/machine communication, as such this paper presents empirical evidence towards the types, and effects of context on human communication; with a view to how this will affect human/machine communication. Results from the study have shown that recognition improvements in humans are over 30% higher, when exposed to contextual information compared to either limited or no context. Moreover, further results show that the first two-thirds of a word are of highest importance for human word recognition. These results are discussed in detail, alongside the methodology of the research, with a view to how these results can impact human/machine communication. The presented research shows that not only does context benefit human communication, but that the use of context has the potential to considerably improve the performance of speech recognition systems.**

***Index Terms*—Machine Learning, Contextual Information Processing, Natural Language Processing**

I. Introduction

Human communication is normally supported by additional contextual information that aids understanding. Context itself, as described by the Oxford dictionary, is: ”The circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood.” [1].

A lack of, or misinterpretation of context often results in a misunderstanding, which then tends to negatively impact human communication and at the least affect its efficiency. In contrast to human communication, which relies quite extensively on context, Machine Learning (ML) algorithms only support contextual information processing to a limited extent within their design.

This research aims to explore how context affects human communication, with a view to how ML can learn to process context in a similar manner to humans. Tools that use ML, such as Natural Language Processing (NLP) and Speech Recognition (SR), are expected to benefit from processing context, specifically by improve recognition rates. Furthermore, an investigation into the background of context, in both computational and human research has also been undertaken, which helps to support the design and rationale of the contextual approach.

II. Background

While context has previously been used in computing fields, the types and variety of context are limited, often relying purely on location data. However, humans often use a wide range of contextual information, from location all the way to a persons emotional state. Clearly, the full range of context is not always used in human communication, but having a wider range than what is often used in computing, is beneficial for communication. Support for this can be seen in fields such as psychology, which is where the background review starts before moving into how context is currently used in computing.

A. Context from a Psychology perspective

An area in which humans benefit from context, can be seen in the case of memory recall, something known as context dependent memory. Context dependent memory is the process by which human memories become linked with external stimuli, for instance location, environment or a number of other factors. Research has shown that humans remember details much better when the context is similar or familiar to when a memory or task is first performed [2], [13].

Furthermore, research has also shown that exposing someone to background noise, which is connected to a performed task, can improve a persons proficiency in that task and aid memory recall [12]. Considering that contextual information aids memory recall, it shows that humans not only derive/use context in communication, but also in how they store information within the brain, which indicates its overall significance.

Moreover, further research indicates that adults with a limited reading ability, often use contextual information present in surrounding text, to make inferences as to the meaning of sections they cannot understand [5]. An ability that has implications for spoken languages as well, for when a person is not fully aware of words being spoken (such as not hearing something correctly) they can often still understand the intention, if not always the full content, of the communication.

Clearly, contextual information, such as sounds, situations and communication style, allow humans to infer certain details that do not need to be explicitly stated, such as the mood of a person. Consequently, context must also improves a person’s ability to make judgments by inferring important information, thus ”filling in the blanks” that can occur during communication. Utilisation of this form of contextual processing could have a plethora of benefits of NLP and ML implementations.

B. Context within Computing

While contextual information has already been used in the field of context aware computing, the types and variety of contexts are limited. For example, some tools keep track of an individual's location to provide them with personal information, such as phone calls, by routing this information/call to the nearest available work-station [8], [14].

Other means of obtaining contextual information can be derived from e.g. accelerometer or light sensor output, as currently found in mobile devices [9]. In this case, the obtained information is used to determine the correct orientation of images on a screen or to automatically adjust brightness of the screen, and so the contextual information is used to adapt processes that relate to the context/mode of operation. Alternatively, accelerometers have also been used as an authentication method for mobile devices [7].

Considering that context is all about the situation someone/something is presented with, it is this type of input that would be beneficial to ML. This could be obtained through images and/or sounds and has been demonstrated through guidebooks that recognise artwork, to help tailor a museum visitor's experience [11], but also in using video to improve speech recognition rates [3].

Events are also a useful form of providing contextual input, for example, patents filed by [6] demonstrate how a car crash can be used to trigger responses from a cars' control systems. While context in sentences has also been used in Hidden Markov Models and Deep Neural Networks to aid speech recognition by determining individual utterances [4], [10].

As seen in the previous examples, very few systems utilise multiple types contextual information, such as time, location, mood, calendar information or the wide range of contextual information stored average smart phone, to identify or benefit from a context. Mostly, they use one or two contexts for a specific purpose, but not as a factor in system learning or training, where the connections between communication and context could be beneficial, such as in SR or NLP. Yet, there is the potential that by having these contextual cues that both SR and NLP recognition rates could be improved, due to an improved understanding of the related contexts.

III. Methodology

To evaluate the effects of context within human communication, a survey was created, and then made available to participants through an online survey tool. The survey was conducted online so that it could attract a diverse audience worldwide, ensuring results are as representative of the population as possible. Having results that are representative in this fashion, helps to ensure that any potential demographic bias would be identified, should it exist.

The survey consists of several questions covering: context within text, context within audio and how incomplete audio is best understood. The design of this survey has taken inspiration, in part from [15].

A. Text with Context

The first set of tests aims to explore the difference in recognition accuracy for sentences with one or multiple types of context. In these tests, the context with multiple contextual types was text (due to the large amount of context present in even simple sentences), and the singular type of context was audio. The text included statements relating to a situation, person(s) or actions, while the audio only related to a locational/situational clue e.g. background nosie.

The text based tests comprise of three questions, which are sentences with two omissions. These questions each consist of four multiple-choice answers. For example, one of the sentences, is: ”Mark walked down to his local [omitted word]. He was hoping to pick up some food for his [omitted word] tonight.”. Options to fill the space are: ”Bank-Dog”, ”Council-Meeting”, ”Shop-Dinner” or ”Beach-Party”. Questions are designed so that only one of the available choices could be considered correct. The use of two-part answers, provided extra context, so that the participants could further match contexts (in addition to those found in the question) thus aiding the selection of the correct answer. The audio part of this test is explained further in the next subsection.

B. Context for Auditory Information

The tests in this section, considered the importance of context within audio, as well as comparing the difference between audio with and without context. Each time there were 3 questions, where one word was obscured by noise. In the cases with context, background sounds were added to provide a contextual clue, while the cases without context had no such sounds. In practice, each audio sample contained a single sentence: ”The next departure will be at [omitted word] 4”, where the omitted word is either ”platform” or ”gate” relating directly to a train station or airport setting.

Participants were asked at the start to indicate how familiar they were with the various modes of transport, thus reducing potential bias based on previous knowledge of the presented context. During the test, the participants had the freedom to play the sounds at their leisure, while they were asked to choose from the following options: ”Gate”, ”Platform” and ”Not Sure” always appearing in that particular order.

C. Word Part Significance

Further audio related tests aimed to explore the importance of each part of a single word. For which, several words were split into 3 parts, and these were respectively obscured at either the front, middle or back. Overall, six words were used in this test, of which there was one ”pair” per word-part. Each word was only used once to ensure that memory or prediction would not influence the results.

Each pair consisted of one word with two syllables e.g. ”Mother”, and one with three syllables e.g. ”Daffodil”. Words are provided as audio samples and there was no order in

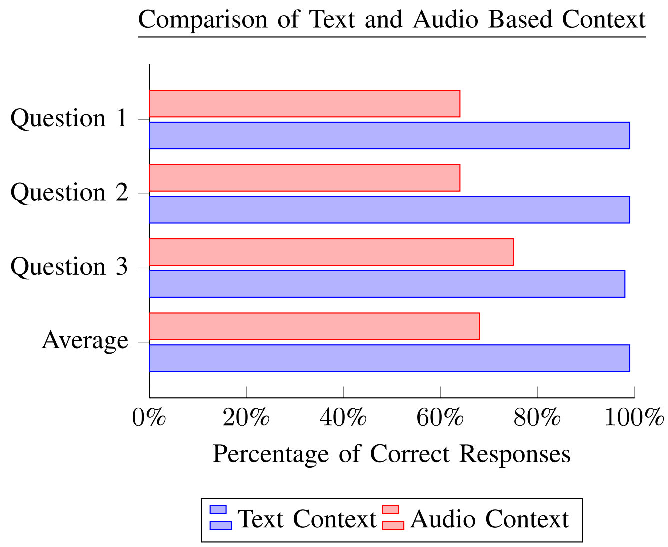

Fig. 1. Comparison between audio with context (one type of context) and text with context (multiple types of context)

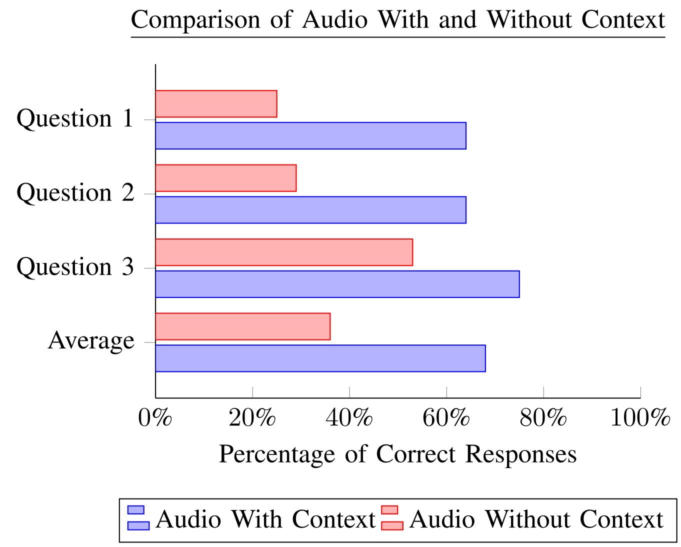

Fig. 2. Comparison between questions with/without context

the way they were presented to the participant with regards to number of syllables and/or part of word that had been obscured. Participants were asked to provide answers in a free text box. Which, as this led to a variety of results, the results were graded as 1 for a correct answer, 0.5 for a partially correct answer (e.g. sounds similar or has correct part of word as hearable present in the solution) and 0 for an incorrect answer.

IV. Results

Results from the study are presented below, and where appropriate subjected to a Student t-test. The demographic section details the demographic spread of participants in the survey.

A. Demographics

The study comprised of 80 participants, with a gender split of: 38 females, 28 males, 13 unidentified and 1 non-binary. The average age was 44, and ranged from 22 to 79. At the start of the tests, each participant was asked for details pertaining their hearing ability through an online hearing test. Further introductory questions asked for their familiarity using various modes of transport (Train, Plane, Boat, Bike, Bus, Taxi), their first language, nationality, profession and ethnicity. These question were asked so that comparisons could be run with specific demographics focuses, to ensure no particular bias affects the results. The related comparisons found no particular bias in the results relating to demographic information.

B. Text and Audio with Context

Figure 1 details a comparison between text and audio based context. In this example the text based context has multiple types of context, while the audio based context only has one type of context.

As seen in Figure 1 the text with context has an accuracy of nearly 100%, illustrating that the text based context has all the contextual information needed to decipher the correct answer.

On the contrary, audio with context averages to only about 68%, which shows that limiting the amount of context available, significantly lowers the response accuracy. Clearly, having a wider variety and amount of contextual information improves the overall responses, in this case by over 30%.

A noteworthy point, is that although text based context had a higher accuracy than audio based context, the main difference (other than information type) was the additional context in the text compared to audio. Thus audio with the same level of context, as the text based context, is likely to improve by similar amounts when provided alongside more types of context.

Additionally, when subjected to a t-test the results show a probability distribution of 5e-10, which indicates that both data sets are statistically different, or in other words, that more context results in better recognition.

C. Audio with and without Context

The comparison between audio with context and audio without context, is shown in Figure 2. Each having three questions. The figure details the accuracy of the participant responses.

The results in Figure 2 illustrates that accuracy is considerably higher for the questions with context, than those without. In this case the audio with context shows improvements of over 30%. Clearly the lack of context severally weakens the accuracy of a participant's response, while adding just one type of context can cause considerable improvements.

Interestingly, there is a slight peak at question three for both audio with and without context. The reason for this seems to lie in the most popular choice (due to familiarity) for an answer, which was "Platform". "Platform" was the correct

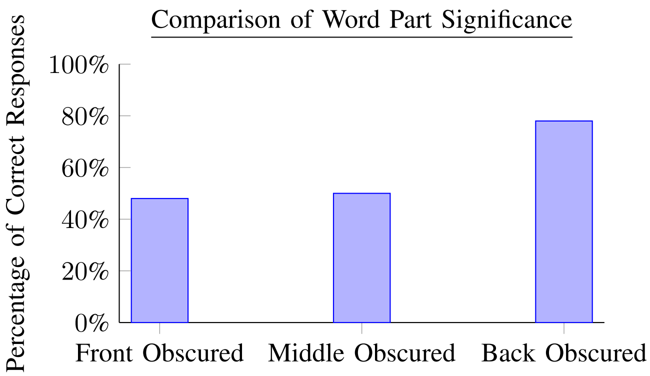

Fig. 3. Comparison of which part of a word is most significant

answer for question three and incorrect for one and two, thus given the peak it would seem that those who selected the most familiar option went with ”Platform” including on question three. Thus explaining this peak.

A combined analysis of audio with/without context, subjected to a t-test shows a probability score of 3e-10 once again indicating that both data sets are statistically different, and consequently supporting the hypothesis that context improves human recognition rates.

D. Word Part Significance

Figure 3 shows the results of the word part significance test. This test aimed to identify the parts of a word which are most important for understanding the whole word. In the test, words were obscured at the front, middle or back one-third.

From the results, it is clear that the highest recognition accuracy of the words, is achieved when the back/latter part of the word is obscured. On the contrary, recognition accuracy is poorer when the front or middle part are obscured. This indicates that the most significant part of a word lies in the first parts, and that if a person misses the last part of a word, inferring the correct word is more viable, while this applies much less for missing the other parts of the word. This can also be noticed in practice where in various languages and/or dialects the endings of words may be ”swallowed”/not clearly pronounced, which often does not significantly affect the understanding.

V. The Benefits of Context for Machine Learning

The results of the survey, as presented in this paper, support the fact that context is a highly valuable and effective way to improve the understanding of information. While some ML algorithms have implemented contextual information processing, there are very few that do. It is expected that further use of context for ML will provide better understanding of learned data, which should allow for improved inference when data is not clear, thanks to the connection between the data and its related contexts.

This will require ML algorithms to be able to deal with more data e.g. the data pertaining to the current task and the contextual data linked to this task. This contextual information could include information about time-frames, background sounds and locational information, among others. In the context of SR and NLP, the contextual data potentially from background sounds, could help with reducing possible misunderstandings and improve overall recognition, based on human responses in this paper.

Although one needs to keep in mind that humans can switch context very quickly and the tools should be able to perform similarly. ML algorithms may not match this speed immediately. However, by training an ML algorithm with its task specific training data, and related contextual data at the same time; the overall learning that the ML possesses should intrinsically link the context with the task specific data. Thus the response speed of a contextual ML tool should be similar to non-contextual ML tool, although training time may be longer. Yet, the resultant improvements in recognition accuracy should counter the increased training times.

VI. Applicability to Natural Language Processing and Speech Recognition

Contextual information is an important part of human communication, not only indicated in the results from this survey, but also from the background research undertaken as part of this paper. Since SR and NLP both deal with, albeit somewhat differently, human communication, it stands to reason that context can improve recognition accuracy.

As mentioned in the previous section, it is likely that an ML tool will need to be trained with task specific and contextual data. So that it to be able to recognise a context and the relationships between that context and the potential words that are being spoken. It is expected this contextual information can be gathered either form background noise, or even from the information already available on a modern smartphone.

Once a ”contextual model” exists that can link contexts to known phrases and words, it is expected that this will allow NLP/SR systems to make better inferences about what has been spoken by a user, hopefully limiting the frustration that can often come with these technologies. The expectation is that when a word or set of words that are linked to a context appear, or a context is recognised, then the SR/NLP tool will have an extra factor to strengthen its certainty of a certain word or phrase being correct. It will also allow ”odd” words that do not fit the context to be reconsidered, thus avoiding an SR/NLP tool offering an unusual word, rather than a more reasonable choice.

VII. Conclusion

This study has confirmed that context has a positive effect on human communication and that generally more context leads to better understanding. It also shows that the first two-thirds of a word is more significant for inferring the whole word, in contrary to the last one-third.

There were three levels of context looked at in this study, multiple types, a singular type and none. In review the multiple types of context performed best, with 30% improvements over just a singular types of context. While a singular type of context also performs 30% better than no context at all.

Overall, there are various benefits of using more contextual information in NLP and SR as well as ML, although it is likely to result in increased data requirements and the need for longer learning times. However, it is expected that a larger amount of contextual information should result in better overall recognition, which is anticipated to offset the expected higher learning times and data requirements of the approach. Future research may also be able to lower these requirements when contextual processing is better understood and can be more efficiently linked into ML training.

VIII. Further Work

Further work will look to explore the types of context that are beneficial to SR/NLP systems and how a context can be identified and used to improve their respective recognition rates. To test this theory a simple SR system will be developed to recognise/process both speech and context, with a view to training a relationship between the speech and context using ML. It is expected that by training the SR system in this fashion, future SR accuracy will be improved upon the identification and use of contextual information.

IX. Acknowledgments

The authors wish to thank Josh Davis for his help with distributing the context survey.

References

[1] Oxford University Press, 2017, Oxford Living Dictionary.

[2] D. R. Godden and A. D. Baddeley, "Context-dependent memory in two natural environments: On land and underwater," *British Journal of psychology*, pp. 325–331, 1975.

[3] A. Gupta, Y. Miao, L. Neves, and F. Metze, "Visual features for context-aware speech recognition," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 5020–5024.

[4] S. Kanthak and H. Ney, "Context-dependent acoustic modeling using graphemes for large vocabulary speech recognition," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2002, pp. I–845.

[5] S. Ng, B. R. Payne, E. A. Stine-Morrow, and K. D. Federmeier, "How struggling adult readers use contextual information when comprehending speech: Evidence from event-related potentials," *International Journal of Psychophysiology*, pp. 1 – 9, 2018.

[6] J. Palmer, M. Freitas, D. A. Deninger, D. Forney, S. Sljivar, A. Vaidya, and J. Griswold, "Systems and method to trigger vehicle events based on contextual information," Feb. 1 2018, uS Patent App. 15/727,227.

[7] A. Primo, V. V. Phoha, R. Kumar, and A. Serwadda, "Context-aware active authentication using smartphone accelerometer measurements," in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2014, pp. 98–105.

[8] B. N. Schilit, D. M. Hilbert, and J. Trevor, "Context-aware communication," *IEEE Wireless Communications*, pp. 46–54, 2002.

[9] A. Schmidt, M. Beigl, and H.-W. Gellersen, "There is more to context than location," *Computers & Graphics*, pp. 893–901, 1999.

[10] F. Seide, G. Li, X. Chen, and D. Yu, "Feature engineering in context-dependent deep neural networks for conversational speech transcription," in *2011 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2011, pp. 24–29.

[11] L. Seidenari, C. Baecchi, T. Uricchio, A. Ferracani, M. Bertini, and A. D. Bimbo, "Deep artwork detection and retrieval for automatic context-aware audio guides," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, p. 35, 2017.

[12] S. M. Smith, "Background music and context-dependent memory," *The American Journal of Psychology*, pp. 591–603, 1985.

[13] S. M. Smith and E. Vela, "Environmental context-dependent memory: A review and meta-analysis," *Psychonomic bulletin & review*, pp. 203–220, 2001.

[14] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems (TOIS)*, pp. 91–102, 1992.

[15] R. M. Warren and R. P. Warren, *Auditory illusions and confusions*. WH Freeman, 1970.