





# Düzce University Journal of Science & Technology

Research Article

## Evaluation of Vocal Communication in a Robot Collective

 Mehmet Dinçer Erbaş<sup>a,\*</sup>,  İsmail Hakkı Parlak<sup>a</sup>

<sup>a</sup> Computer Engineering Department, Engineering Faculty, Bolu Abant İzzet Baysal University, Bolu, Turkey

\* Corresponding author's e-mail address: dincer.erbasm@ibu.edu.tr

DOI: 10.29130/dubited.688255

### ABSTRACT

In this research, we attempt to design a model in which multiple robots communicate with an artificial proto-language whose symbols are vocally encoded letters of the Morse alphabet. We have shown that, as the robots have limited sensing and acting abilities, the communicated symbols of the proto-language differentiates from their original versions due to copying errors. We check the effects of two distinct environmental factors, namely the positional distance between the robots and the amount of noise in the environment. It is shown that both of these factors affect, in different ways, how accurately the presented proto-language can be accurately transmitted by the robots.

**Keywords:** Vocal communication, robot learning, multi-robot systems.

## Bir Robot Kolektifinde Ses ile Haberleşmenin Değerlendirilmesi

### ÖZET

Bu araştırmada, birden fazla robotun, sembolleri Mors alfabesinin sesli olarak kodlanmış harfleri olan yapay bir öncül-dil vasıtasıyla iletişim kurduğu bir model tasarlanmıştır. Robotların sınırlı algılama ve eyleyici yeteneklerine sahip olduklarından, kullanılan öncül-dilin iletilen sembollerinin kopyalama hataları nedeniyle orijinal sürümlerinden farklılaştıkları gösterilmiştir. İki ayrı çevresel faktörün, robotlar arasındaki konumsal mesafe ve ortamdaki gürültü miktarının, kopyalama üzerine etkileri incelenmiştir. Bu faktörlerin her ikisinin sunulan öncül-dilin robotlar tarafından ne kadar doğru bir şekilde iletilebileceğini etkilediği gösterilmiştir.

**Anahtar Kelimeler:** Ses ile haberleşme, robot öğrenmesi, çok-robotlu sistemler.

# **I. INTRODUCTION**

Language, that allows humans to communicate by means of a collection of symbols that can be vocally transmitted, is one of the key cognitive abilities that make humans unique. For long years, many researchers have attempted to examine, model and finally understand how and why languages can emerge and shape the cognitive activities of human beings. These efforts mostly stemmed from the need to distinguish human languages from other types of communication systems. For this purpose, in his seminal work, Hockett [1] declared 13 design features that are shared by all natural languages. Based on these features, he was able to comparatively discuss the differences between human languages and the methods that are used by animals to communicate. The first five of these features, namely *vocal auditory channel*, *broadcast transmission and directional reception*, *rapid fading*, *interchangeability* and *total feedback* model the physical properties of the vocal channel that is used as a medium for transferring the symbols of the language. The next three features, *specialization*, *semanticity*, and *arbitrariness* enhance the expressiveness of the communication system. The next two features, *discreteness* and *traditional transmission* explain the mechanisms in which linguistic units are transferred between individuals. Hockett claimed that the last three features, namely *displacement*, *productivity* and *duality of patterning* make the human language unique among all communication systems. According to the displacement feature, humans are unique in the sense that they are able to talk about things that are remote in space or time. The last two features are particularly significant as they make the human languages open systems in which new linguistic utterances can emerge and be included in the repertoire of the language in time. Based on the productivity feature, Hockett stated that new linguistic units that had never been used before can be understood by other users of the language and these units can gradually become a part of the language. Duality of patterning explains how distinct building blocks, namely morphemes of the language, can be arranged and combined in different ways to form new linguistic utterances.

Among the features that were presented by Hockett, the last two, in effect, make the human languages dynamic systems that can progress in time. As they have a significant effect in making the human languages dynamic systems, the productivity and duality of patterning principles have been widely examined by some researchers. These researchers claimed that the languages are complex dynamic systems that can adapt to changing conditions under the influence of individual learning, cultural adaptation and biological characteristics [2], [3]. In order to examine the adaptive functionality of languages, a number of different methods have been utilized, including computer simulations [4], [5] and experiments with human participants [6], [7]. In these researches, to model the dynamics of linguistic items of a language, the participants of the experiments interacted with each other through a process of social coordination [2], [3], [4] or iterated learning [8], [9]. The iterated learning [10] method particularly have received much attention in recent years. This method simulates a process of repeated learning and usage during which an agent learns a specific linguistic item through observing the demonstration of the item by another agent that shares the same environment and who learned it in the same way. Experiments on iterated learning have modeled adaptive communication systems through which agents can communicate effectively in laboratory conditions.

In this research, we attempt to design an experiment setting in which multiple robots communicate through an artificial proto-language whose symbols are vocally encoded letters of the Morse alphabet. We have shown that, as the robots have limited sensing and acting abilities, the communicated symbols of the proto-language differentiates from their original versions due to copying errors. We check the effects of two distinct environmental factors, namely the positional distance between the robots and the amount of noise in the environment. It is shown that both of these factors influence how accurately the presented proto-language can be transmitted by the robots. Finally, we discuss the utility of our model and mention some possible future work.

The article proceeds as follows: Section II mentions some related works. Section III presents the vocal communication algorithm and quality of vocal communication function that are used in the experiments.

In section IV, we present the real robot experiments on vocal communication and discuss their results. Finally, section V concludes the paper and mentions some further future work.

## **II. RELATED WORK**

An issue that is worth mentioning about most of the researches on the evolution of language is that in the models that were developed, the participants learned artificial symbolic proto-languages through which they can communicate over the visual channel. The symbols that were transferred between participants consisted of written strings or graphical images that undergo variations during multiple cycles of acquisition and reproduction. As the learning process was not modeled over the vocal channel, the dynamics and physical properties of sound emission/reception did not have any effect on the adaptation of linguistic units. As stated above, some of Hockett's designed features applies to the dynamics of the vocal channel so it is definitely worth examining the adaptation of language over a vocal communication task.

There have been a few attempts to model the adaptation of artificial proto-languages over a vocal channel. For instance, Verhoef [11] examined the development of a simple artificial whistled language through which a number of human participants tried to learn specific sound combinations that could be reproduced with slide whistle. Starting with an initial set of whistles, each participant listened and then reproduced the heard set of sound combinations which were then used as the training items for the next participant. In this way, Verhoef was able to model the adaptation of a proto-language over a vocal channel. In the experiments, the participants were forced to reproduce a specific number of unique sound combinations so that the number of distinct combinations was constant during all learning cycles. With these settings, as multiple learning cycles proceeded, Verhoef was able to observe the development of a more structured and easy to learn whistled proto-language. As the learned set of whistles were passed from participant to participant through a process of repeated learning and usage, the symbols of the language became better suited to the perceptual abilities of the participants so that they could be more accurately reproduced and acquired. In this research, as the participants were forced to remember the heard sound combinations, under such memory constraints, they were able to progressively agree on the structure of the learned symbols so that the adapted symbols could be transferred with high accuracy.

In another research, Zuidema and de Boer [4] utilized computer simulations to model the emergence of novel linguistic units in a group of simulated agents that communicated through an artificial proto-language. The symbols of the proto-language consisted of signals which were represented as trajectories that were derived from specific recorded acoustic data. By introducing random noise to the transfer of symbols between agents, they were able to observe the emergence of combinatorial structure in the adapted symbols as novel signals could be formed by recombination of basic building blocks.

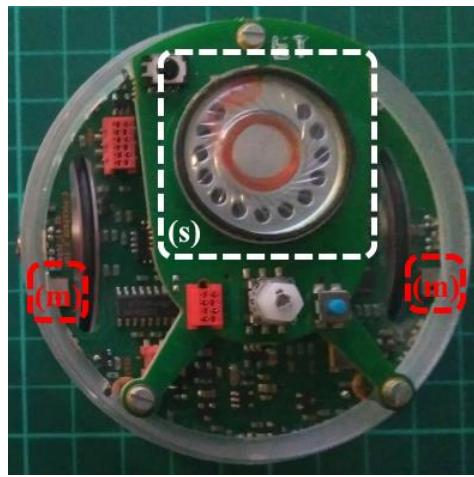
One of the key mechanisms in the research on the adaptation of language is that there should be a source of variance in the learning process so that the learned linguistic units can adapt to uncertainties in the environment. In experiments with human participants, the existing biases of the participants or memory constraints provide the necessary variances. In experiments with computer simulations, the needed variances are artificially introduced to the system. In this respect, real robot experiments offer an interesting alternative platform. Similar to the natural systems, robots have limited perceptual abilities and they can partly sense their environment through a noisy medium. In addition, they can be programmed to learn visually or vocally from other robots or humans. Based on these observations, there have been some research on language adaptation in recent years in which real robots were used as participants in the experiments [12]. In these researches, robots visually learned from each other and the adaptation of the learned linguistic units were examined. However, there has been no research that attempts to model the adaptation of a language that can be learned through the vocal channel on real robots. In this research, we attempt to design an experiment setting in which multiple robots vocally communicate through an artificial proto-language. In contrast to the related work presented in this section, our model uses real robots whose less-than-perfect perceptual and motor system, introduce the

necessary variations to the vocal learning process. In this way, we are able to examine the dynamics of vocal communication in a controlled environment.

### III. METHOD

#### A. EXPERIMENT SETUP

We use e-puck miniature robots [13] in our embedded vocal communication experiments. The e-puck robots are mobile, compact and they are equipped with several sensors and actuators, including three microphones and a speaker as shown in figure 1, so the robot has the necessary apparatus for on-board vocal communication. The maximal acquisition speed of its microphones is 33 kHz.



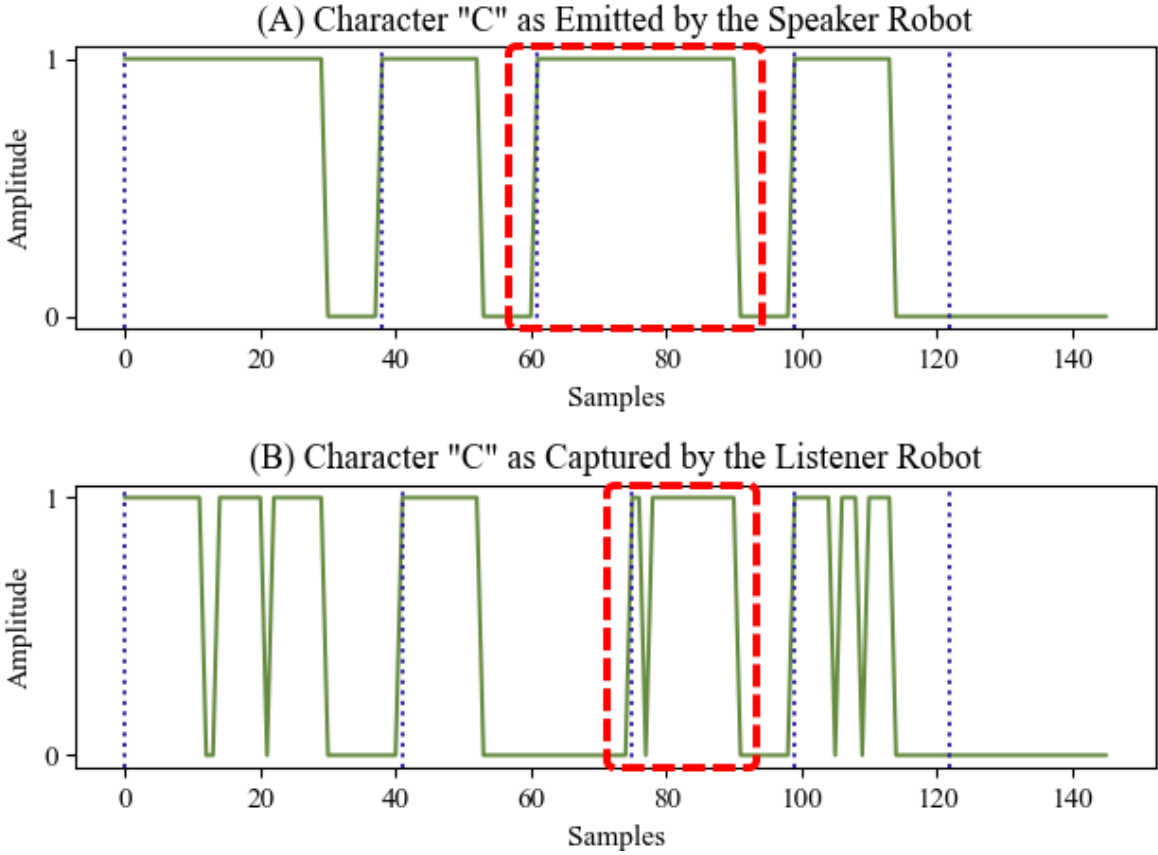
*Figure 1. An e-puck robot. The speaker of the robot is shown in the white square marked with (s); the microphones of the robot are shown in smaller red squares marked with (m). The third microphone is located under the speaker's base plate.*

The e-puck robot can play embedded audio files through their speakers. It comes with an embedded audio file comprised of baby sounds and noises. For this study, the robot's embedded audio files are overwritten with an 880Hz sine wave which has an 8kHz sampling frequency. In order to eliminate audio distortion in the speakers, we reduce the level of the audio files by a factor of 0.5. Any louder audio signal causes distortion when it is played through robot's speakers.

During the experiments, two robots vocally communicate through their on-board sensors. At the start of each experiment run, one of the robots is declared as the "speaker" while the other robot is declared as the "listener". Then, they attempt to communicate by generating different combinations of short and long beeps that mimic the letters of the Morse code. Each letter of the Morse code is encoded into 5 intervals, separated with silent durations, during which the speaker robot emits short/long beeps or it stays silent. For instance, figure 2(a) shows the encoded version of the letter "C" of the Morse code. As can be seen, in order to communicate the letter "C", the speaker robot emits a long beep in the first interval, a short beep in the second interval, a long beep in the third interval, a short beep in the fourth interval and finally it stays silent in the fifth interval.

The speaker robot, when it starts communicating, selects a letter of the Morse code, converts it into a sequence of short/long beeps or silent intervals, and emits the corresponding audio signals. As the speaker speaks, the listener robot captures the audio signals that it received from its microphone and converts the received signals into a sequence of short/long beeps and silent intervals. After speaker robot completes speaking, robots switch their roles, i.e., listener robot becomes speaker and speaker robot becomes listener. In this way, we are able to model the vocal communication of the letters of Morse code in a group of robots.

However, as the robot’s sensors and actuators are not perfect, we observe copying errors in the reproduced audio signals. For instance, in figure 2(a) an example audio signal, which corresponds to the letter “C” of the Morse code, is shown. Figure 2(b) shows the audio signal that is reproduced by the listener robot when it listens to the execution of the letter “C” by the speaker robot. As can be seen, there are some copying errors in the reproduced audio signal. The actual letter “C” of the Morse code consists



**Figure 2.** (a) The encoding of the letter “C” of the Morse code. (b) Its reproduced copy. The erroneous part of the induced copy is shown in dashed rectangle.

of a long beep, followed by a short beep, a long beep, a short beep and a silent interval. However, in this example, the listener robot perceives the emitted signal as a long beep followed by three short beeps and a silent interval. Thus, introducing an error where the third beep should have been a long beep instead of a short beep. The erroneous segment is shown in Figure 2 within dashed rectangles. In order to qualitatively compare an executed audio signal with its reproduced copy, we need a quality of vocal communication metric. For this purpose, we take each audio signal as a string of short/long beeps or silent intervals and compare the normalized Levenshtein distance, in terms of number of insertions, deletions or substitutions, between the original emitted signal and its reproduced copy. Therefore, the quality of vocal communication for a specific audio signal S and its reproduced copy C is calculated as follows:

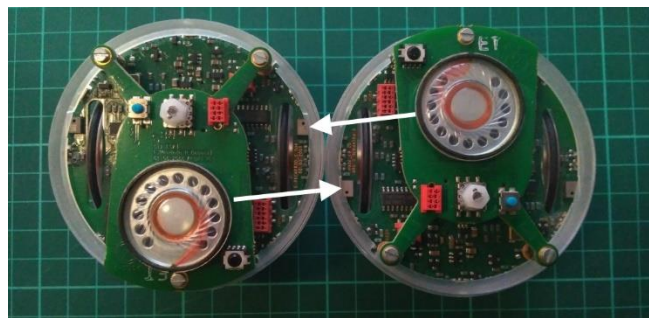
$$\text{Quality}(S, C) = 1 - \frac{\text{Levenshtein}(S,C)}{\text{Max}(\text{len}(S),\text{len}(C))} \tag{1}$$

Based on this metric, the quality of communication for the audio signal shown in figure 2 is calculated as 0.8.

## IV. EXPERIMENTS

The main objective of this research is to examine the dynamics of vocal communication on a group of robots that can communicate with each other the vocal channel. As stated above, there are a number of features of the vocal-auditory channel that make vocal communication particularly more complex than visual communication. For instance, through the vocal channel we observe broadcast transmission and directional reception in which the audio signals that are generated by the speaker rapidly fade away. Therefore, the spatial distance between the speaker and the listener is a significant metric that effects the fidelity of communication over the vocal channel. In a collective group, whose members are mobile, the agents may need to communicate through different spatial distances.

In order to examine the effects of spatial distances on the communication of two mobile robots, we design experiments in which two robots, the speaker and the listener, vocally communicate by transferring audio signals that mimic the letters of the Morse code at different distances. We test the quality of communication for three different distance settings, as shown in figure 2. In the first set of experiments, the robots are placed right next to each other so that the distance between the speaker of the speaker robot and the microphone of the listener robot is minimal. In the second set of experiments, the robots are placed 5 cm away from each other. Finally, in the third set of experiments, they are placed 10 cm away from each other. During the experiments, at each of the specific distance between robots, the speaker robot selects one of the letters of the Morse code, as shown in table 1, converts it to a sequence of short and long beeps and emits the corresponding signal. The listener robot captures the signal, converts it to a sequence of short and long beeps and saves the reproduced sequence in its memory. When the experiments are completed, we compare the original letters that are emitted by the speaker with their reproduced copies, by using the quality of vocal communication function that is presented in the previous section.



(a)



(b)



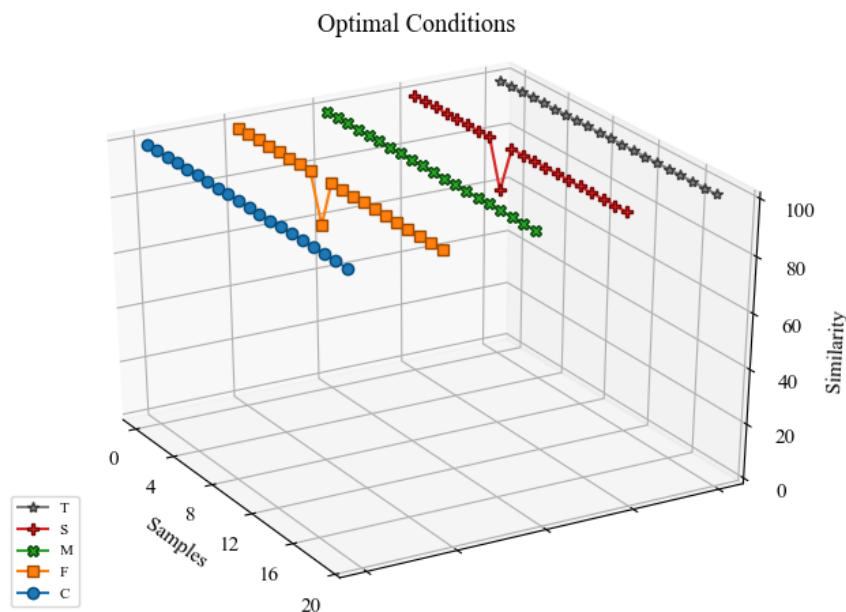
(c)

**Figure 3.** (a) In the first set of experiments, the spatial distance between the robots is minimal (b) In the second set of experiments, the distance between robots is set to 5 cm. (c) In the second set of experiments, the distance between robots is set to 10 cm.

**Table 1.** The letters of the Morse code that are used in experiments. The specific letters are chosen to have distinct combinations of short/long beeps and silent intervals.

Character	Morse Code Representation (Beeps)
C	Long, Short, Long, Short, Silent
F	Short, Short, Long, Short, Silent
M	Long, Long, Silent, Silent, Silent
S	Short, Short, Short, Silent, Silent
T	Long, Silent, Silent, Silent, Silent

Figure 4 shows results of the first experiment set when the distance between robots is minimal. As can be seen, the listener robot is highly successful in capturing the emitted letters. There are a few occasional errors, due to noisy reception of the sensors, however, in most runs, the listener robot was able to determine the exact sequence of signals. In these settings, the listener robot achieves 99.6% quality of communication value (std: 2.81) and over 100 communication runs, in 98 of them, it is able to detect the correct sequence of short/long beeps and silent intervals.



**Figure 4.** The quality of vocal communication when the robots are placed next to each other.

Figure 5 shows the results of the experiments when the robots are placed 5 cm away. As can be seen, they have a slightly lower rate of quality of vocal communication. In these settings, the listener robot achieves 97.5% quality of communication value (std: 7.59) and over 100 communication runs, in 89 of them, it is able to detect the correct sequence of short/long beeps and silent intervals.

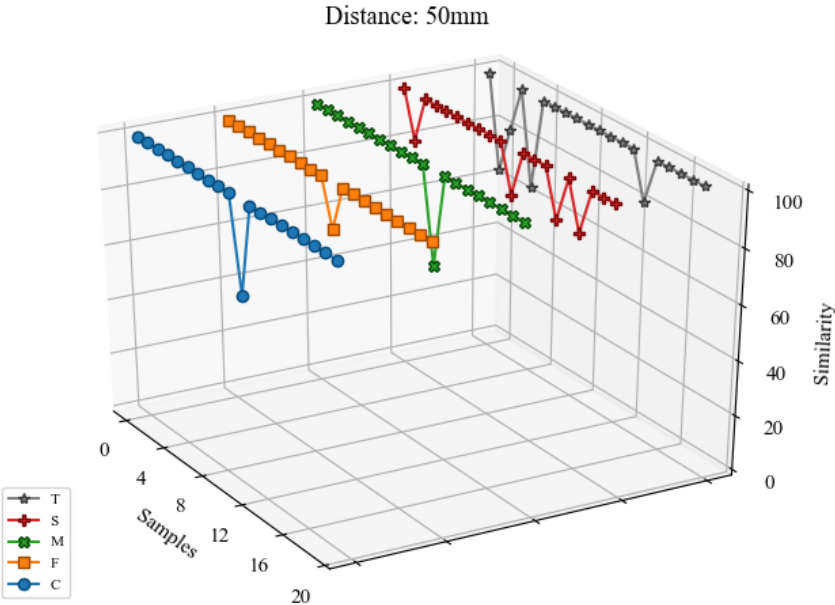


Figure 5. The quality of vocal communication when the robots are placed 50 mm apart.

Finally, figure 6 shows the results of the experiments when the robots are placed 10 cm away. As can be seen, the listener robot has a lower performance in capturing the emitted signals. In these settings, the listener robot can achieve 84.83% quality of communication value (std: 17.65) and over 100 communication runs, in only 44 of them, it is able to detect the correct sequence of short/long beeps and silent intervals.

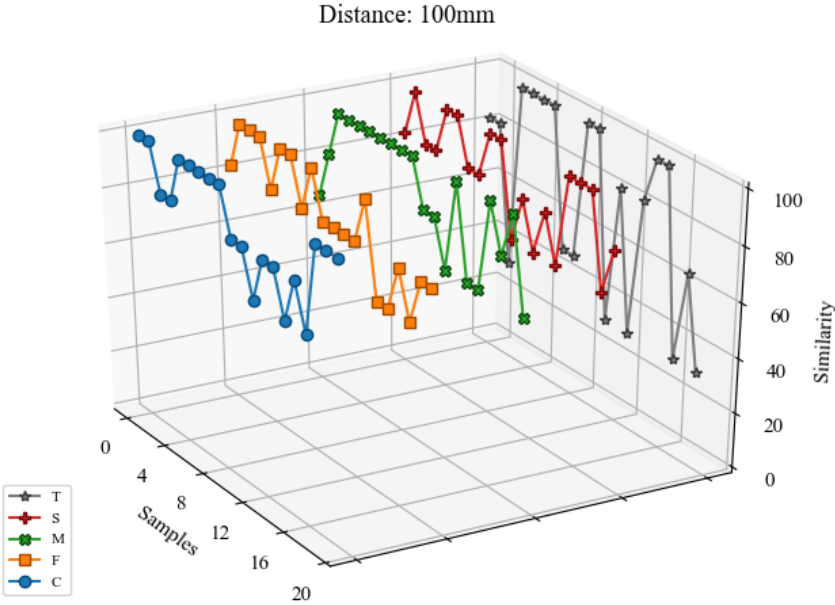


Figure 6. The quality of vocal communication when the robots are placed 100 mm apart.

The audio signals that are emitted by the speaker robot can be categorized based on the number of distinct sound emissions. For instance, the letter T, which consists of a single long beep is relatively a



simple letter, while the letter C, which consists of a combination of 4 distinct sound emissions (long, short, long, short beeps) can be seen as a relatively complex letter. Based on this categorization, when we compare the quality of communication of distinct letters, it can be observed that the listener robot achieves a higher quality of communication value when it imitates simple letters in comparison with complex letters. The highest quality of communication value is achieved when the listener robot imitates letter T (mean: 89.17, std: 12.42) while the lowest quality of communication value is achieved when the listener robot imitates the letter C (mean: 73.33, std: 28.30). A pair-wise ttest between the copies of letter T and the copies of letter C reveals that the difference between the two is statistically significant. Therefore, we can deduce that, if the sound signals that are transmitted vocally are relatively complex in terms of number of distinct sound emissions, an increase in the spatial distance between the robots has a relatively higher negative effect on the quality of communication.

One of the key mechanisms of vocal communication is that, due to the broadcast transmission feature of the auditory channel, if there are multiple audio signals in the same environment that are emitted from different sources, they may mix up so that the listener receives a combination of the signals. Human brain is able to concentrate its attention on a particular audio stimuli by filtering out other audio signals that it receives. This ability is called *the cocktail-party effect*, and the process in which humans can achieve the filtering has been widely studied by many researchers [14], [15].

In a collective group, whose members are mobile, agents may need to communicate in noisy environments in which multiple agent-to-agent dialogues may take place simultaneously. In order to test the effects of noise in the environment, we design experiments in which two robots communicate with each other, as explained above, in an environment that includes artificially generated white noise. For this purpose, the robots are placed next to each other, as shown in figure 3 (a) and we introduce an additional sound source that emits 40 dB of ambient noise. Once again, the speaker robot selects different letters of the Morse code, converts the selected letters into a sequence of short/long beeps and silent intervals while the listener attempts to capture the audio signals.

Figure 7 shows results of the experiment set with 40 dB ambient noise. As can be seen, the ambient noise effects the reception of the listener as it achieves a slightly lower performance compared to the results shown in figure 4. In these settings, the listener robot achieves 96% quality of communication value (std: 7.49) and over 100 communication runs, in 77 of them, it is able to detect the correct sequence of short/long beeps and silent intervals.

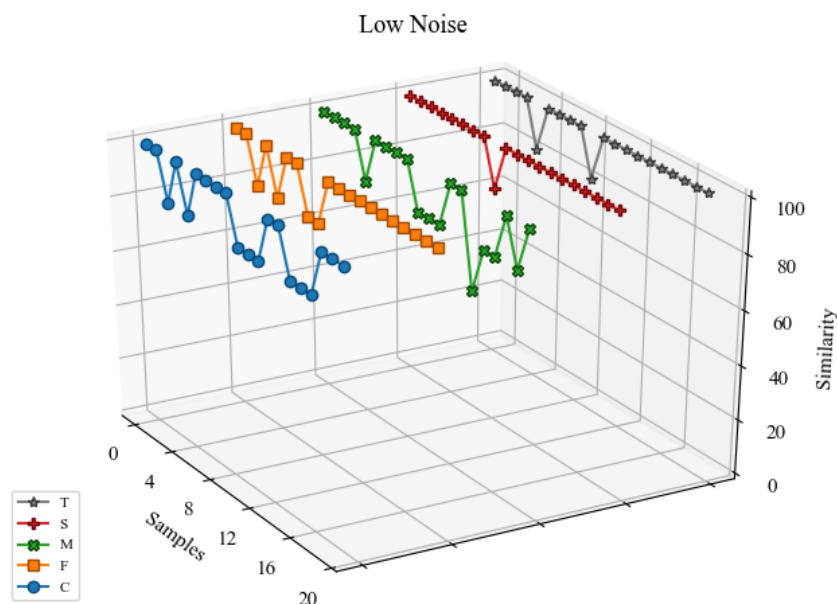
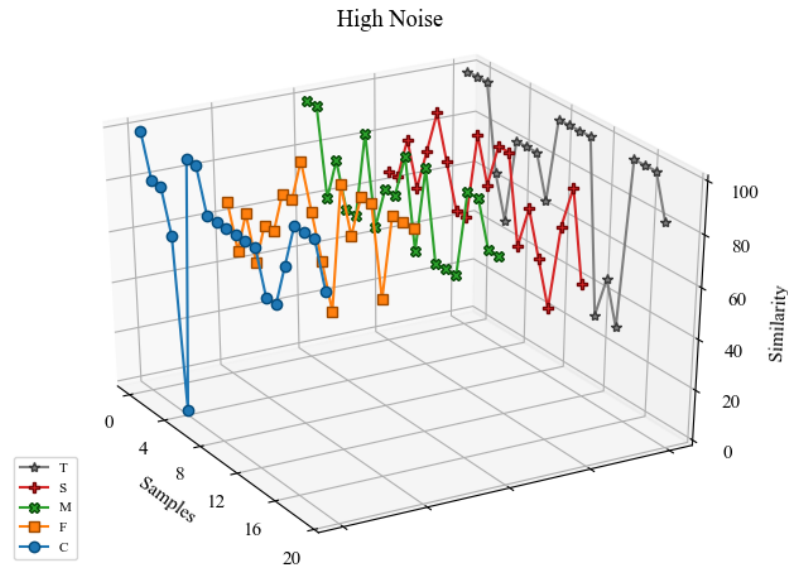


Figure 7. The quality of vocal communication with 40 dB ambient noise.

Finally, to further observe the effects of noise in the environment, we increase the level of ambient noise to 50 dB and repeat the same experiments. Figure 8 shows the results of these experiments. As can be seen, the listener robot has a much lower rate of quality of vocal communication. In these settings, the listener robot achieves 80.83% quality of communication value (std: 18.76) and over 100 communication runs, in only 34 of them, it is able to detect the correct sequence of short/long beeps and silent intervals.



*Figure 8. The quality of vocal communication with 50 dB ambient noise.*

When we examine the quality of communication for each letter individually, it can be observed that the listener robot achieves highly similar quality values for each of the 5 letters (mean quality value for the letters C, F, M, S, T are 81.67 (std: 23.51), 79.17 (std: 15.17), 82.5 (std: 14.78), 79.17 (std: 18.63), 81.67 (std: 22.23), respectively). In contrast to the results shown in figure 6, there is no statistically significant difference in the quality of communication values between the copies of complex and simple letters. Based on this observation, we can deduce that the errors due to noisy environmental conditions manifest themselves as random errors that affect each transmitted signal in a similar rate. On the other hand, the errors induced by the spatial distance between the communicating robots affect the transmission of complex signals in a high rate which make them harder to be learned. In other words, we observe that the errors due to rapid fading feature of the vocal channel introduce different levels of variations on the transmitted signals.

## V. CONCLUSION

In this research, we attempted to model the process of vocal communication in a group of robots who could transmit and receive a sequence of audio signals that mimic the letters of the Morse code. Vocal imitation on the robots was embodied in the sense that all the signal emission/reception activities were executed on-board of the robots. As the sensors and actuators of the robots are not perfect, we observed copying errors while signals were transmitted between the robots. Furthermore, we devised a quality of vocal communication metric that calculates the fidelity of learning through vocal imitation. It was shown that, due to copying errors between robots, the transmitted audio signals differentiated from their originals.

We tested the effects of two distinct factors, namely the spatial distance between the robots and the level of ambient noise in the environment. It was shown that both factors affected how accurately the letters of the Morse code could be transmitted between the robots. The errors due to noisy environmental conditions caused random variations on the transmitted signal while the errors due to spatial distance between the robots favored accurate transfer of simple signals compared to complex signals. By using the embodied vocal imitation framework that is introduced in this paper, it should be possible to examine the adaptation of transferred audio signals during multiple cycles of iterated learning. In this way, we may be able to model the adaptive process during which a proto-language that can be vocally transmitted between robots gets better suited to the environmental factors, so that as time passes, it can be transmitted with higher accuracy.

We tested our model on two robots in which one of them was declared as the speaker and the other was declared as the listener. With a larger group of robots in which multiple robots change roles and communicate at different locations of the environment, we can observe further adaptation of the proto-language that is used by the robots. For this purpose, we plan to design more complex communication protocols that would allow us to examine the adaptation of a vocally transmitted proto-languages in a robot swarm.

## **V. REFERENCES**

- [1] C. Hockett, "The origin of speech," *Scientific American*, vol. 203, pp. 88-111, 1960.
- [2] L. Steels, "The synthetic modeling of language origins," *Evolution of Communication*, vol. 1, no. 1 pp. 1-34, 1997.
- [3] L. Steels, "Human language is a culturally evolving system," *Psychonomic Bulletin & Review*, vol. 24, no. 1, pp. 190-193, 2017.
- [4] B. de Boer, "Self-organization in vowel systems," *Journal of Phonetics*, vol. 28, no. 4, pp. 441-465, 2000.
- [5] W. Zuidema, and B. de Boer, "The evolution of combinatorial phonology," *Journal of Phonetics*, vol. 37, no. 2, pp. 125-144, 2009.
- [6] B. Galantucci, "An experimental study of the emergence of human communication systems," *Cognitive science*, vol. 29, no. 5, pp. 737-767, 2005.
- [7] H. Cornish, R. Dale, S. Kirby, and M. H. Christiansen, "Sequence memory constraints give rise to language-like structure through iterated learning," *PloS one*, vol. 12, no.1, pp. 1-18, 2017.
- [8] B. Chazelle, and C. Wang, "Iterated learning in dynamic social networks," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 979-1006, 2019.
- [9] V. DeCastro-Arrazola, and S. Kirby, "The emergence of verse templates through iterated learning," *Journal of Language Evolution*, vol. 4, no. 1, pp. 28-43, 2019.
- [10] S. Kirby, T. Griffiths, and K. Smith, "Iterated learning and the evolution of language," *Current Opinion in Neurobiology*, vol. 28, pp. 108-114, 2014.
- [11] T. Verhoef, "The origins of duality of patterning in artificial whistled languages," *Language and Cognition*, vol. 4, no. 4, pp. 357-380, 2012.

- [12] L. Steels, "Modeling the cultural evolution of language," *Physics of Life Reviews*, vol. 8, no. 4, pp. 339-356, 2011.
- [13] C. M. Cianci, X. Raemy, J. Pugh, and A. Martinoli, "Communication in a swarm of miniature robots: The e-puck as an educational tool for swarm robotics," in *International Workshop on Swarm Robotics*, 1st ed., Berlin, Germany: Springer, 2006, pp. 103-115.
- [14] J. E. Peelle, "Speech comprehension: Stimulating discussions at a cocktail party," *Current Biology*, vol. 28, no.2, pp. 68-70, 2018.
- [15] S. Getzmann, J. Jasny, and M. Falkenstein, "Switching of auditory attention in cocktail-party listening: ERP evidence of cueing effects in younger and older adults," *Brain and Cognition*, vol. 111, pp. 1-12, 2017.