NAIST-IS-MT1551021

Master's Thesis

Analyses of Communication between Human and Dog

Rina Ouchi

March 6, 2017

Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Rina Ouchi

Thesis Committee:

Professor Kazushi Ikeda	(Supervisor)
Professor Shigehiko Kanaya	(Co-supervisor)
Associate Professor Takatomi Kubo	(Co-supervisor)
Professor Kazuyuki Samejima	(Co-supervisor)

イヌーヒト間のコミュニケーションの解析*

大内 里菜

内容梗概

コミュニケーションはヒトを含む生き物が社会生活を送る上で重要な役割を担っている．同種間のみならず，異種間でのコミュニケーションも知られているが，イヌは進化の過程を通じてヒトとの優れた異種間コミュニケーション能力を獲得したと考えられている．イヌがそのような素晴らしいコミュニケーション能力を獲得しているにも関わらず，ヒトはしばしばイヌと上手にコミュニケーションが取れない場合がある．上手にコミュニケーションが取れない場合，噛みつきや無駄吠えのような家庭犬の問題行動や，災害救助犬のパフォーマンスの低下に繋がることがある．これらの問題行動を解決するためには，ヒトがイヌの内的状態を的確に推定することが求められる．イヌの専門家は，それまでに培ったイヌに関する知識やイヌとのふれあいから得られる経験のような専門知を用いて，イヌの内的状態を上手に推定することができる．しかし，その専門知が具体的にどのようなものであるか，未だ科学的に十分に明らかにされていない．そこで本研究では，イヌのふるまいの情報からイヌの内的状態を推定するための二種類の手法を提案し，調査する．一つ目の手法では，イヌに取り付けた加速度計から得られた加速度データを用いてイヌの内的状態の推定を試みる．イヌの加速度データを使用する理由は，加速度データにはイヌの情動に関わる情報を含むイヌの動きが記録されていると考えられるためである．二つ目の手法では，イヌの専門家の視線データを用いる．その理由は，視線データは，イヌの専門家がイヌとヒトのコミュニケーションを捉える時にどこに注目しているかを反映すると考えられるためである．そして，この視線データから，我々はイヌ―ヒト間のコミュニケーションを観察する時にどこに着目すべきかを学ぶことができると考えられる．上記の理由から，我々はこれらの手法を用いた次の二つの実験を行った．一つ目の加速度データを用いた実験では，快・不快・中立の単純な三種の情動状態におけるイヌの加速度データを計測した．得られた加速度データから，その際の情動状態の識

*奈良先端科学技術大学院大学 情報科学研究科 情報科学専攻 修士論文, NAIST-IS-MTMT1551021, March 6, 2017.

別が可能かどうか機械学習の手法を用いて検証を行った．視線計測を用いた実験では，イヌの専門家と非専門家それぞれを被験者として，ドッグトレーニングのビデオを視聴している際の視線データを計測し，専門家と非専門家の間で比較を行った．前者の実験の結果，イヌの個体毎の加速度データを使用した場合，全てのイヌの加速度データを使用した場合における 3 種の情動状態の識別を行ったところ，それぞれの 92%，71%の精度で識別が可能であった．後者の結果より，イヌの専門家はイヌ—ヒト間のコミュニケーションを視聴している時に独自の視点を持つことが示された．これらの結果から，イヌは情動状態に応じた独自の行動パターンを持つこと，そして専門家はイヌの様子から内的状態に関する情報を読み取ろうとしている可能性があることが示唆された．

キーワード

コミュニケーション，イヌ，加速度データ，視線，専門知

Analyses of Communication between Human and Dog*

Rina Ouchi

Abstract

Communication plays an important role in social life for animals including humans. It is believed that dogs have acquired excellent communication skills with humans through the process of evolution. Although they have such skills, people sometimes cannot communicate with them well. Miscommunication leads to issues such as problematic behavior of pet dogs or lowered performance of rescue dogs. In order to solve these problems, it is required to infer the inner states of dogs appropriately. Dog experts can infer dogs' states using their knowledge or experience as one of their expertise. However the expertise has not been well understood. In this study, we examine how to infer inner states of the dogs from their behavior. The first approach is using data measured with accelerometer equipped on a dog and machine learning technique to decode the emotion of dogs. We assumed that dog's motions should have information related to its emotions. The second approach is using gaze behavior of the dog expert because the gaze points might reflect their attention to dogs. From their gaze behaviors, we can learn what we should observe during human-dog communication. Thus we conducted two experiments. In the experiment for the first approach, the acceleration data of dogs were measured under the three types of simple emotional conditions, positive, negative and neutral. In the experiment for the second approach, the gaze data of subjects including one dog expert are measured with eye tracker during free viewing of a video of dog training.

These results suggest that dogs have actions reflecting their emotion and dog experts can catch the signals of dogs that may reflect the inner states.

*Master's Thesis, Graduate School of Information Science,
Nara Institute of Science and Technology, NAIST-IS-MT1551021, March 6, 2017.

Keywords:

communication, dog, acceleration data, gaze, expertise

Contents

List of Figures

1 Introduction

Communication is important for humans. We communicate with others using the various methods. For example, gaze, voice, gesture and so on. These communication skills are used to live and cooperate with others. However, these communication skills are used not only by humans but also by other animals. Especially, some non-human animals are getting attention because they are similar animals to humans. They are known that they have good communication skills [1].

In a recent study, it is revealed that dogs also have excellent communication skills with humans. For example, the dogs can understand pointing gesture by human [2] and can follow human's gaze point [3]. Moreover, another previous research has suggested that dogs' communication skills are result of evolution [4]. In this evolution, a hormone named oxytocin and gaze play an important role. Oxytocin is known that it is used to promote an attachment behavior between mother and infant. It is also known that oxytocin effects between conspecific animals. However, the previous study shows oxytocin effects between allospecific animals. For example, when a dog and a human gaze each other, their oxytocin concentrations increase. On the other hand, when their oxytocin concentrations increase, they gaze each other more frequently. In this meaning, oxytocin may affect dogs' communication with humans like mother-infant communication. Therefore, dogs have acquired the skill during social life with humans. Thanks to this communication skill, dogs have been good partner for humans.

However, we sometimes cannot communicate with the dogs well. First example is the pet dogs do problem behaviors such as biting or barking. Second example is lowering performance of the rescue dogs. The one of causes of these cases are that some people cannot infer the inner states of dogs appropriately. To infer the inner states of the dogs are not easy for non-expert people by using the dogs' appearance because it is need a lot of knowledge and experience to infer it [5]. However what information is needed to infer the inner states is unknown.

In this study, we developed an inference method of the inner states to make communication between dogs and humans smooth because sometimes non-expert people cannot do appropriate behavior to the dogs.

Here we describe two ideas to infer the inner states of dogs.

First idea is using information of the dogs except its appearance to infer the states objectively. This is described in chapter 2.

Second idea is analyzing information of dog expert who can infer the inner state appropriately. In this study, we extract feature of expert's gaze point to examine what is important information for communication with dogs. This is described in chapter 3.

2 Estimation of emotions using an accelerometer

2.1 Introduction

Emotion, which is critical for working and learning, affects a person's performance both mentally and physically. For example, a strong motivation in rehabilitation improves performance [6]. This effect can even be applied to dogs. Some handlers of rescue dogs say rescue dogs show a lower performance during the task of finding injured people after disasters when they have negative emotions. In such cases, the handlers have to find a substitute. Thus, they need to recognize the animal's emotional state to maximize the performances of rescue dogs.

An emotional states in dogs are estimated from physiological data since emotion induces changes such as oxytocin secretion, cortisol concentration, heart rate and heart rate variability [7–10]. However, since physiological data are difficult to measure and easily corrupted by environmental noises in practical situations, simpler sensors are more convenient and preferable.

In this chapter, we proposed a method to estimate emotional states using acceleration data. The acceleration data are easy to measure robustly and rich enough to estimate the action states of humans, horses and dogs [11–13]. Since emotional states also affect actions, they would be estimated from acceleration data.

2.2 Materials and Methods

Our estimation system consists of three parts: measurement, preprocessing, and classification. The second extracts effective features for classification and the last

Figure 2.1: Procedure of the experiment.

discriminates a given condition into one state, positive, negative, or neutral. The classifier was trained by the labeled data from experiments (see data collection). The data were also used to evaluate our system's performance using the cross validation [14].

2.2.1 Data Collection

The acceleration data in this study were collected in the experiment in [10] and [15]. The experiment was approved by the ethical committee of Azabu University.

Subjects

39 healthy dogs and their owners in Azabu University were recruited (Age, mean $\pm$ SE = 4.56 $\pm$ 0.56 years).

Procedure

In this experiment, three types of condition were prepared, neutral, positive, negative.

In the beginning, the owner stayed with his/her dog for five minutes (neutral condition). Next, the owner called the dog's name and gently petted it for five minutes (positive condition) and then returned to the neutral condition for five minutes. Next the owner departed from the experiment room and left the dog alone for five minutes (negative condition) before returning to the room (Fig. 2.1). However, to avoid the order effect, the order of presentation of positive and negative condition was varied for each dog.

Figure 2.2: Accelerometer.

Measurement

The three-dimensional acceleration data were collected by a three-dimensional accelerometer (TSND 121, ATR Promotions, Japan) at a sampling frequency of 50 Hz (Fig. 2.2). ECG data were simultaneously collected by a sensor (TSEMG01, ATR Promotions, Japan). Three electrodes connected to the sensor were placed on the chest along the sternum and the sensor was attached to the dog's back, between the scapulae. However the ECG data were not used in the following analysis.

2.2.2 Preprocessing

To make features that are robust to the device-orientation, we calculated the acceleration's magnitude, which is invariant against device direction [11]. The features of our system are the powers of the short-time Fourier transform (STFT) with a window-size of 10.24 sec. and a shift-time of 1.28 sec.

2.2.3 Classification

Our system can employ any classifier in machine learning and signal processing [16]. In our study, we tested Support Vector Machine (SVM) [17] and Random Forest [18] since both of them are well-known and high generalization performance classifiers in the machine learning community.

SVM in our system was implemented using the R package 'e1071' [19]. The package provides an interface to libsvm which is a fast and easy-to-use implementation of the most popular SVM formulations and includes the most common kernels (linear, polynomial, RBF, and sigmoid). In our study, we used linear and RBF kernels. Since the problem was a multiclass classification, we employed one-versus-one strategy.

Random forest in our system was implemented using the R package 'randomForest' [20]. Random Forest is an ensemble learning method for both classification and regression that constructs a number of decision trees at training time and chooses its output class by voting.

The SVM and Random Forest parameters were chosen by grid search to maximize their performances using the cross validation [14].

2.2.4 Evaluation

We used 70% of the collected data for training and the remainder to evaluate the estimation ability.

The estimation was done in two ways: Estimation within a subject and estimation over all subjects. In the former, each classifier was trained using the dataset from one dog and evaluated using the same dataset. In the latter, the data from all the dogs were mixed for training and evaluation.

2.3 Results

Nine of 39 dogs were excluded from the analysis because three dogs did not lie down within an hour, three failed in recording their RRI due to equipment trouble, and three showed unusual behavior during the positive condition [10].

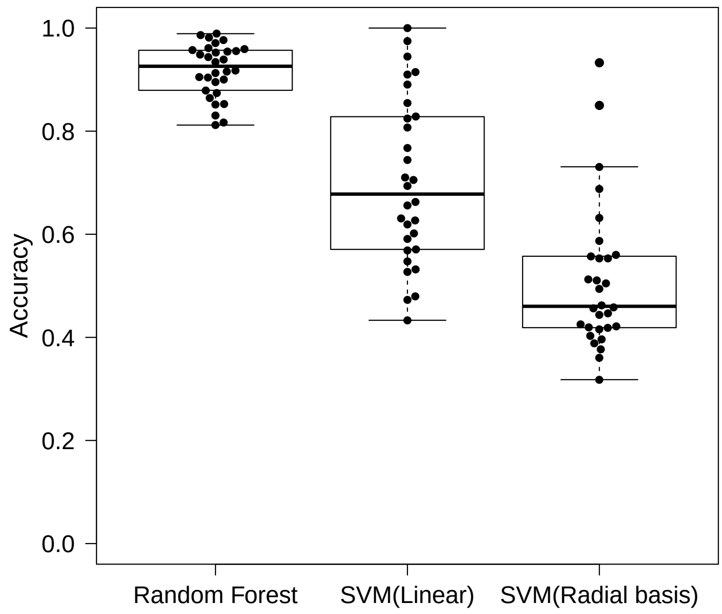
In both estimation cases within a subject and over subjects, Random Forest showed better accuracy (92% and 71%, respectively) than SVM with the linear kernel (70% and 41%, respectively) or the radial basis kernel (50% and 41%, respectively). in classifying positive, negative, and neutral emotions (Fig. 2.3).

2.4 Discussion

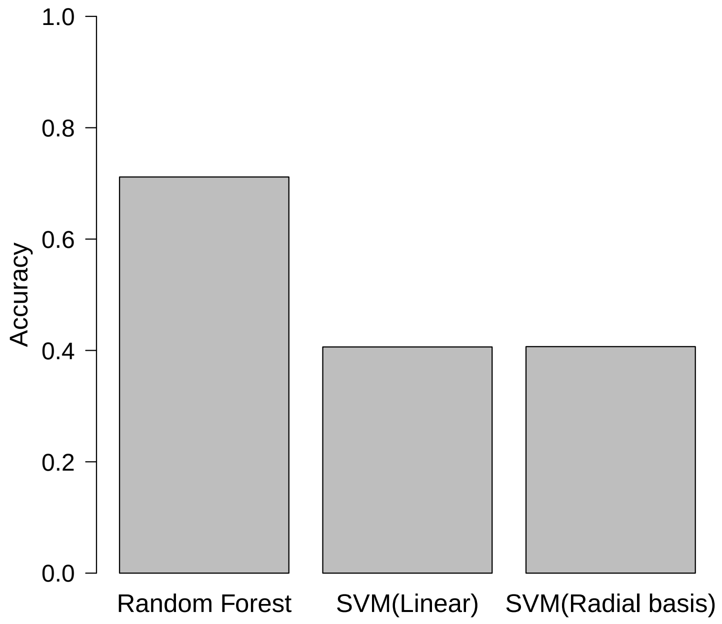
Although the accuracy of 71% over all subjects of Random Forest was not high, because the rescue dogs were registered in advance and were collected data to make a tailored classifier. Hence, the classifier is practical for rescue dogs since the accuracy of 92% within a dog was surprisingly high.

Compared to the estimation accuracy of 92% of Random Forest, those of SVMs with linear and radial basis kernels were much lower. This tendency was seen in other applications [21].

We proposed a method for emotional state estimation from acceleration data. The method, which consisted of the STFT based preprocessing and a Random Forest classifier, classified each dog's emotional state with 92% accuracy on average when trained with the dog's data. The accuracy is high enough for practical use in rescue dogs or pet dogs.



(a) estimation within a subject



(b) estimation over subjects

Figure 2.3: Estimation accuracy of the proposed system.

3 Identification of features in behaviors for communication

3.1 Introduction

Non-experts cannot infer inner states of the dogs from dogs' behavior information because dogs' behaviors do not correspond one to one with their inner states. On the other hand, it is believed that dog experts can infer such inner states. Also, techniques to infer it have been introduced in some books including textbook for dog trainers [5].

However, this expertise has not been well understood from the scientific viewpoint. On the other hand, some other kinds of expertise, e.g., art [22] or dance [23] come to be treated scientifically. To examine these expertise, researchers have measured behavioral data such as gaze data of experts and novices, and compared them. Although, for these expertise, inferring others' internal states are not necessary, dog experts have to infer the dogs' emotional states. Regardless of this difference, gaze behavior might be still a hint to understand how the experts infer the internal states of dogs from their behavior.

In this chapter, to study the features of gaze behavior of dog-experts, we measured gaze data of dog expert and novices and compared them as a pilot study.

3.2 Materials and Methods

3.2.1 Subjects

One dog expert and six novices participated in our experiments. The dog expert is a student of the Azabu University. The novices does not have academic knowledge about dogs.

3.2.2 Stimulus

Stimulus was a dog training video by a dog trainer (available from on the web [23], uploaded by K9-1.com). It was selected by a dog researcher who is a collaborator of our research. The length of this video was 285 seconds. This video shows a scene of dog training where one dog trainer instructs a puppy dog or an adult dog to lie on a mat using food or gesture. The puppy dog was under the process of training. In contrast, the adult dog had already done learning the training and can perform it well.

3.2.3 Apparatus

For the eye-tracking equipment, we used Tobii TX300 Eye Tracker (Tobii Technology AB, Sweden), which consists of infrared cameras. The equipment was located in front of a 23-inch monitor, which displayed the stimulus at a screen resolution of 1920 × 1080 pixels. Each participant sat at a distance of 65 cm from the monitor. The visual angle of the stimulus was approximately 31 degrees squared. Vertical and horizontal coordinates of gaze locations of the participant's eye were recorded at a sampling rate of 120 Hz. Stimulus presentation and eye-tracking were controlled using Matlab (Mathworks) toolboxes, psychtoolbox [25] and Tobii Pro Analytics SDK [26].

3.2.4 Procedure

Participants were instructed to view freely the video displayed as the experimental task. They were informed that displayed video was dog training video. Prior to the recording, calibration was conducted by collecting fixations on target points, where the participant gazed on each of five points located on the four corners and the center on the screen. The video was displayed in our trial, which consisted of a blank period of 3 seconds and a presentation period of 285 seconds. In the blank period, a white cross was displayed at the center of the screen for 3 seconds, and the participants fixated on it. Subsequently, the participants viewed the dog training video, with gaze locations of the eye recorded. We excluded the gaze data during the scenes which do not show any communication between the dog and the trainer (e.g., narration scene or caption scene) from the following analysis.

3.2.5 Data Analysis

We defined the four regions of interest in the video according to the categories of objects (Dog, Trainer, Background, Others). These categories are used to confirm which object the subjects paid attention to. Moreover, we divided these 4 categories into 12 finer categories (Dog: body, face, hip, legs. Trainer: bag, body, face, hands, legs. Background: mat, others. Others). To quantify the frequency of gaze, we counted the number of gazing for each subjects and categories, for every one second. And then, we calculated averaged frequency of the novices and Z score of the experts for each category.

And then, we calculated average of novices data and Z score of each category.

3.3 Results

In the 4 rough categories, expert watched the dog in the video more often than novices ($Z = 2.77$, $p < 0.01$), and novices watched the dog as often as any other objects (Background: $Z = -2.98$, $p < 0.01$, Others $Z = -2.02$, $p < 0.05$) in the video (Figure. 3.1).

In the 12 finer categories, the expert watched more often dog's face ($Z = 3.10$, $p < 0.01$) and dog's hip including its tail ($Z = 5.12$, $p < 0.01$) than novices (Figure 3.2).

In contrast, novices often watched Background ($Z = -3.40$, $p < 0.01$), Others ($Z = -2.02$, $p < 0.05$), trainer's hands ($Z = -2.95$, $p < 0.01$), and dog's legs ($Z = -2.08$, $p < 0.05$).

3.4 Discussion

From this result, the expert watched the dog more often than the novices and focused on its face and hip including tail. It is known that these parts reflect

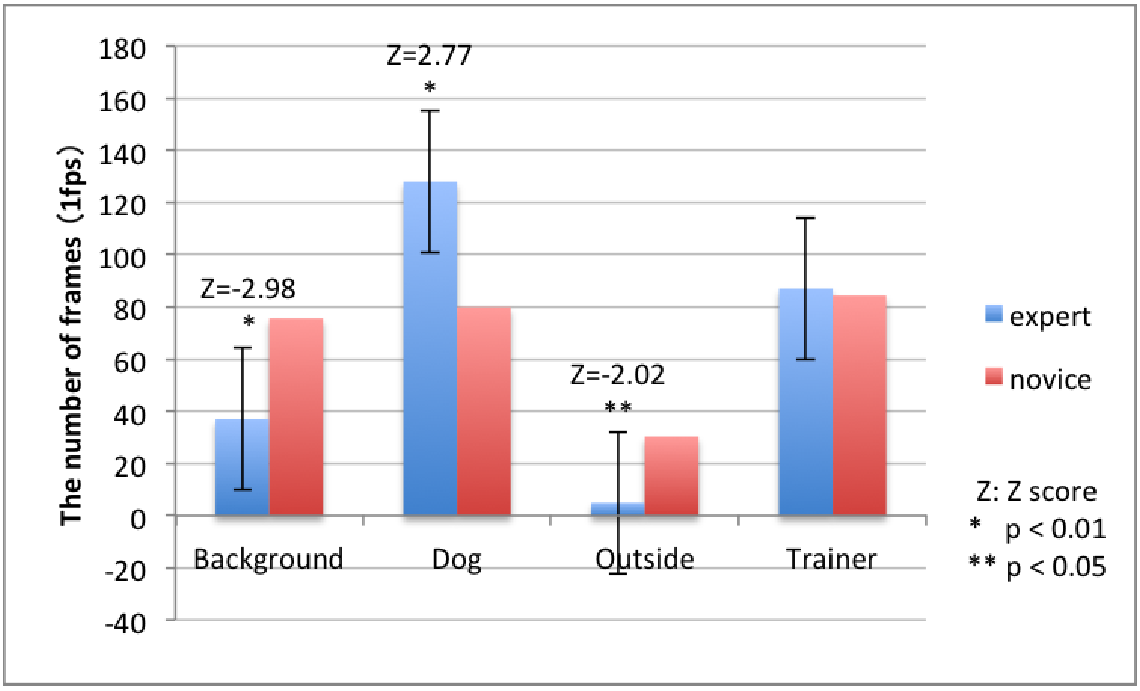


Figure 3.1: The total numbers for each of the categories (rough, four categories). Error bar: standard deviation.

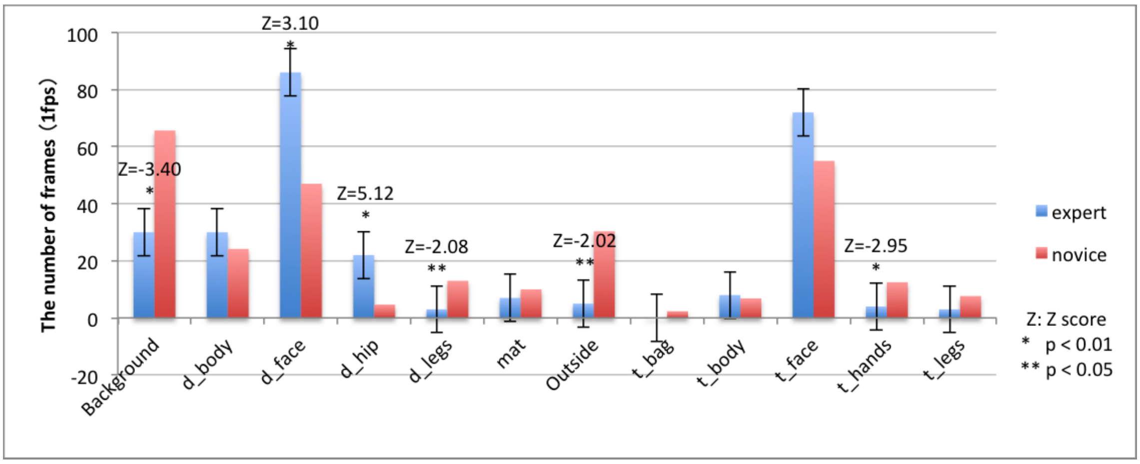


Figure 3.2: The total numbers for each of the categories (fine, twelve categories) "d_" means a part of the dog. "t_" means a part of the dog trainer. Error bar: standard deviation.

emotion of the dogs.

In contrast, novices watched trainer's hands and dog's legs. These objects were moving in the video frequently. Therefore, It may mean novices are affected by just movements in the video. To confirm this, it is needed to analyze data considering quantification of movement.

We assumed that expert might have unique gaze point during watching communication between the dog and the trainer. Furthermore, the difference might show evaluation criterion made by knowledge and experience of expert.

On the other hand, a previous study showed that the dog experts and the novices had same view points during watching still images of an interaction between a pair of dogs or a pair of humans [27]. This study also measured fMRI of dog expert and novices to reveal the difference between them.

Therefore, to examine unique gaze points of dog's expert, it is necessary to compare not only the gaze behavior data but also fMRI data between novices and experts. The contents of the video are trainer/novice train a dog. Also, it is preferable to confirm the difference of gaze points between experts and novices using multiple videos with different skill level of actors (expert and novice).

4 Conclusion

Communication with dogs is important for humans to live and cooperate with them. To make communication between humans and dogs smooth, we need to infer the inner states of dogs appropriately.

In this study, as the method to realize that, first, we tried to infer simple emotional states of the dogs by combining dogs' acceleration data and machine learning algorithm. From the result, it is assumed that we can infer the simple states from acceleration data, however this method cannot be used in practical communication scene because this method cannot be generalizable between dogs and needs acceleration data of target dogs with annotation about emotional state in advance. Therefore, we also tried to reveal what is important for dog expert during watching communication between them by comparing feature of gaze points of a dog expert and the novices. From the result, it is suggested that experts have unique gaze points depending on their knowledge and experience. The expert focused on the different points which were dog's face and hip including tail compared with novices' view points. These points are known as parts for emotional expression. In contrast, the novices focused on trainer's hand and dog's legs which moved frequently in the video. It is presumed novices simply reacted the feature of the movement in the video.

As the result of this study, however, we cannot show what is important to communicate for humans with dogs in daily life. To reveal it, we try to study unique view points of the dog experts using new experimental data. During new experiment, we also gave subjects questionnaires. In the questionnaires, we asked them if a dog trainer could judge dogs' states appropriately or if the dog paid attention to the dog trainer, and why the subjects think so, and so on. Therefore, we will keep on investigate what is important for communication by combining gaze data and the results of assessment such as goodness, generality, etc.

Acknowledgements

I could not have written this thesis without the guidance and advice of my advisors, Professor Kazushi Ikeda, Professor Shigehiko Kanaya, Associate Professor Takatomi Kubo, Professor Kazuyuki Samejima and collaborators and the help of my lab member, friends and family.

References

[1] Shinya Yamamoto et al., “ Chimpanzees Help Each Other upon Request, ” PLOS ONE, vol. 4(10), e7416, 2009.

[2] K. Soproni et al., “Dogs’ (Canis familaris) responsiveness to human pointing gestures, ” J Comp Psychol, Vol 116(1), pp. 27-34, 2002.

[3] B. Hare et al., “Communication of Food Location Between Human and Dog (Canis Familiaris), ” Evolution of Communication, Vol 2(1), pp. 137-159, 1998.

[4] M. Nagasawa et al., “ Oxytocin-gaze positive loop and the coevolution of human-dog bonds, ” Science 17, Vol 348, pp. 333-336, 2015.

[5] R. Vibeke, “ ドッグ・トレーナーに必要な「深読み・ 先読み」テクニック, ” 誠文堂新光社, 2011.

[6] Y. Nishimura et al., “ Neural substrates for the motivational regulation of motor recovery after spinal-cord injury, ” PLOS ONE, vol. 6, e24854, 2011.

[7] T. Romero et al., “Oxytocin promotes social bonding in dogs, ” PNAS, vol. 111, pp. 9085-9090, 2014.

[8] B. Beerda et al., “ The use of saliva cortisol, urinary cortisol, and catecholamine measurements for a noninvasive assessment of stress responses in dogs, ” Hormones and Behavior, vol. 30, pp. 272-279, 1996.

[9] M. Zupan et al., “Assessing positive emotional states in dogs using heart rate and heart rate variability, ” Physiology & Behavior, vol. 155, pp. 102-111, 2016.

[10] M. Katayama et al., “Heart rate variability predicts the emotional state in dogs,” Behavioural Processes, vol. 128, pp. 108-112, 2016.

[11] Y. Maruno et al., “Energy-efficient user-state recognition method using wavelet transform and singular value decomposition,” SICE-JCMSI, vol. 8, pp. 86-95, 2014.

[12] R. Thompson et al., “Dancing with horses: automated quality feedback for dressage riders,” Proc. UBICOMP, pp. 325-336, 2015.

[13] L. Gerencser et al., “Identification of behavior in freely moving dogs (Canis familiaris) using inertial sensors,” PLOS ONE, vol. 8, e77814, 2013.

[14] G.C. Cawley and N.L.C. Talbot, “On over-fitting in model selection and subsequent selection bias in performance evaluation,” J. Machine Learning Research, vol. 11, pp. 2079-2107, 2010.

[15] E. Nakahara et al., “Canine Emotional States Assessment with Heart Rate Variability,” Proceedings of APSIPA Annual Summit and Conference, 2016.

[16] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2007.

[17] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273-297, 1995.

[18] T.Ho, “RandomDecisionForests,” Proc. 3rd Int’l Conf. Document Analysis and Recognition,” pp. 278-282, 1995.

[19] D. Meyer et al., “Package ‘e1071’,” Ver. 1.5-11, cran.rproject.org, 2005.

[20] A. Liaw et al., “Classification and Regression by randomForest.” R News 2(3), 18-22, 2002.

[21] R. Mao et al., “Comparative analyses between retained introns and constitutively spliced introns in Arabidopsis thaliana using random forest and support vector machine,” PLOS ONE, vol. 9, e104049, 2014.

[22] N. Koide et al., “Art Expertise Reduces Influence of Visual Salience on Fixation in Viewing Abstract-Paintings,” PLOS ONE, vol. 10.2, e0117696, 2015.

[23] C. Stevens et al., “Perceiving dance: schematic expectations guide experts' scanning of a contemporary dance film,” Journal of Dance Medicine & Science, vol. 14, 2010.

[24] Dog Training by K9-1.com, “How to Train a Dog to Go to a "Place" Mat (K9-1.com),” [https://www.youtube.com/watch?v=qLHAKyd3hJ8] (URL), 2013.

[25] Brainard DH, “The psychophysics toolbox,” Spatial Vision 10: 433-436. doi: 10.1163/156856897X00357 PMID: 9176952, 1997.

[26] Tobii Technology, “Tobii Pro Analytics SDK,” [http://www.tobiipro.com/ja/product-listing/tobii-pro-analytics-sdk/] (URL).

[27] Kujala MV et al., “Dog Experts' Brains Distinguish Socially Relevant Body Postures Similarly in Dogs and Humans,” PLoS ONE, vol. 7.6, e39145, 2012.