International Journal of Informatics Society

09/21 Vol. 13 No.2 ISSN 1883-4566

Informatics Society

Aims and Scope

The purpose of this journal is to provide an open forum to publish high quality research papers in the areas of informatics and related fields to promote the exchange of research ideas, experiences and results.

Informatics is the systematic study of Information and the application of research methods to study Information systems and services. It deals primarily with human aspects of information, such as its qu ality and value as a resource. Informatics also referred to as Information science, studies t he structure, algorithms, behavior, and interactions of natural and a rtificial systems that store, process, access and communicate information. It also develops its own conceptual and theoretical foundations and utilizes foundations developed in other fields. The advent of computers, its ubiquity and ease to use has led to th e study of info rmatics that has computational, cognitive and social aspects, including study of the social impact of information technologies.

The characteristic of informatics' context is amalgamation of technologies. For creating an informatics product, it is necessary to integrate many technologies, such as mathematics, linguistics, engineering and other emerging new fields.

Guest Editor’s Message

Tetsushi Ohki

Guest Editor of Thirty-eighth Issue of International Journal of Informatics Society

We are delighted to have the Thirty-eighth issue of the International Journal of Informatics Society (IJIS) published. This issue includes selected papers from the Fourteenth International Workshop on Informatics (IWIN2020), which was held online, Sept. 10-11, 2020. The workshop was the fourteenth event for the Informatics Society, and was intended to bring together researchers and practitioners to share and exchange their experiences, discuss challenges and present original ideas in all aspects of informatics and computer networks. In the workshop 24 papers were presented in seven technical sessions. The workshop was successfully finished with precious experiences provided to the participants. It highlighted the latest research results in the area of informatics and its applications that include networking, mobile ubiquitous systems, data analytics, business systems, education systems, design methodology, intelligent systems, groupware and social systems.

Each paper submitted IWIN2020 was reviewed in terms of technical content, scientific rigor, novelty, originality and quality of presentation by at least two reviewers. Through those reviews 17 papers were selected for publication candidates of IJIS Journal, and they were further reviewed as a Journal paper. We have three categories of IJIS papers, Regular papers, Industrial papers, and Invited papers, each of which were reviewed from the different points of view. This volume includes five papers among those accepted papers, which have been improved through the workshop discussion and the reviewers’ comments.

We publish the journal in print as well as in an electronic form over the Internet. We hope that the issue would be of interest to many researchers as well as engineers and practitioners over the world.

Tetsushi Ohki is an associate professor at Shizuoka University. He received the BE and ME degrees in electronics and communication engineering from Waseda University, Tokyo, Japan, in 2002 and 2004, respectively, and the Ph.D degree in engineering from Waseda University in 2010. He is currently an associate professor in the Faculty of Informatics, Shizuoka University, Japan. His research interests include biometrics, pattern recognition, information security and privacy. He is a member of ACM, IPSJ, and IEICE.

Regular Paper

Sub-1 GHz Wireless Coexistence of IEEE 802.15.4g and IEEE 802.11ah Using Hybrid CSMA/CA

Yukimasa Nagai†*, Jianlin Guo‡, Takenori Sumi†, Philip Orlik‡, and Hiroshi Mineno*

†Mitsubishi Electric IT R&D Center, Japan
‡Mitsubishi Electric Research Laboratories, USA
* Graduate School of Science and Technology, Shizuoka University, Japan
Nagai.Yukimasa@ds.MitsubishiElectric.co.jp

Abstract - As more and more wireless technologies have been developed to support emerging Internet of Things (IoT) applications, the coexistence of these heterogeneous wireless technologies presents challenges. IEEE 802.15.4g and IEEE 802.11ah are two of such wireless technologies specified for the outdoor IoT applications and designed to operate in the Sub-1 GHz (S1G) frequency band. Due to the constrained spectrum allocation in the S1G band, two types of devices may be forced to coexist, i.e., share frequency spectrum. Therefore, the coexistence of 802.15.4g and 802.11ah should be addressed. To investigate coexistence behavior of these two wireless technologies, we first identify coexistence issues using our developed NS-3 based S1G band coexistence simulator. Simulation results confirm that 802.15.4g performance can significantly degrade due to the 802.11ah interference. Accordingly, we propose a hybrid CSMA/CA mechanism for 802.15.4g to address the identified coexistence issues. We then present several distributed and network assisted methods for 802.15.4g devices to estimate 802.11ah interference severity and switch channel access mode for interference mitigation. The performance analysis shows that the proposed hybrid CSMA/CA can improve 802.15.4g performance without sacrificing 802.11ah performance[1].

Keywords: Wireless coexistence, interference mitigation, hybrid CSMA/CA, Sub-1 GHz band, IEEE 802.15.4g, IEEE 802.11ah

1 INTRODUCTION

The Internet of Things (IoT) applications are rapidly growing. A broad range of wireless technologies have been developed to cater diverse applications. As heterogeneous wireless technologies are emerging, wireless coexistence becomes a critical issue to be addressed. IEEE 802.15.4g [2], marketed as Wi-SUN, operates in the Sub-1 GHz (S1G) frequency band for outdoor IoT applications. IEEE 802.11ah [1], marketed as Wi-Fi HaLow, is the first 802.11 standard designed to operate in the S1G band. The unlicensed spectrum allocation is limited, especially in the S1G band compared with other bands such as 2.4 GHz band. For example, Japan only allocates 7.6 MHz spectrum in 920 MHz band for active radio devices in the standard ARIB STD-T108 (20 mW) [3]. This standard also regulates other passive radio devices to use this spectrum. The constrained spectrum allocation indicates that 802.15.4g devices and 802.11ah devices may be forced to coexist, i.e., share frequency spectrum. In addition, 802.15.4g network and 802.11ah network can have thousands of nodes. Both technologies have communication range of 1000 meters for IoT applications. These features significantly increases the coexistence potential. Therefore, ensuring harmonious coexistence of these two wireless technologies is important.

802.11ah mandates the support of 1 MHz channel, which is much narrower than the 20 MHz channel for conventional 802.11 in 2.4 GHz band. As a result, the existing coexistence technologies designed for 2.4 GHz band may not be suitable for the coexistence of 802.11ah and 802.15.4g in the S1G band, e.g., the cooperative busy tone method proposed in [10] assumes 22 MHz 802.11 channel. Therefore, the coexistence of 802.11ah and 802.15.4g needs to be further investigated. Accordingly, IEEE New Standards Committee and Standard Board formed IEEE 802.19.3 Task Group in December 2018 to develop an IEEE 802 standard for the coexistence of 802.11ah and 802.15.4g in the S1G frequency band [4]. Authors of this paper have been leading this standard development.

J. Guo, et al. propose a prediction based self-transmission control method for 802.11ah to ease its interference impact on 802.15.4g [6]. Y. Liu, et al. introduce α-Fairness energy detection clear channel assessment (ED-CCA) method for 802.11ah to mitigate its interference on 802.15.4g caused by its higher energy detection (ED) threshold [7]. To address the interference caused by the faster backoff of 802.11ah, Y. Liu, et al. also propose a Q-Learning based backoff mechanism for 802.11ah to avoid interfering with 802.15.4g packet transmission process [7]. However, these coexistence technologies improve the performance of 802.15.4g at the expense of 802.11ah. This paper aims to develop coexistence technologies for 802.15.4g to improve 802.15.4g performance without degrading 802.11ah performance. We first evaluate coexistence behavior and identify coexistence issues by using the developed S1G band

[1]This paper is an extended version of our work published in [5]. We extend our previous work by adding distributed and network assisted methods for 802.15.4g devices to estimate 802.11ah interference severity. These methods are indispensable for 802.15.4g devices to assess 802.11ah interference and therefore, switch channel access mode for interference mitigation.

Table 1: The majority of available performance evaluation, and conventional coexistence researches.

Reference	Year	Target System	Band	Objective	Validation Tool
This article and [5]	2020	11ah & 15.4g	Sub-1 GHz	delivery rate and latency at coexistence	ns-3
J. Guo, P. Orlik [6]	2017	11ah & 15.4g	Sub-1 GHz	delivery rate and latency at coexistence	ns-3
Y. Liu, J. Guo et al.[7]	2018	11ah & 15.4g	Sub-1 GHz	delivery rate and latency at coexistence	ns-3
W. Yuan et al. [8]	2010	11b & 15.4	2.4 GHz	throughput	OPNET
E.D.N Ndih et al. [9]	2016	11 & 15.4	2.4 GHz	delivery rate	MATLAB
X. Zhang, et al. [10]	2011	11 & 15.4	2.4 GHz	analytical model, throughput	analytical, ns-2
J.Hou et al. [11]	2009	11 & 15.4	2.4 GHz	delivery rate	experiments
J.W. Chong et al. [12]	2015	11 & 15.4	2.4 GHz	throughput	analytical
B. Badihi et al. [13]	2013	11ah & 15.4	Sub-1 GHz	throughput and energy consumption	OMNeT++
R. Ma et al. [14]	2017	11b & 15.4	2.4 GHz	analytical model, throughput	analytical & unknown simulator

coexistence simulator. We then propose a hybrid carrier sense multiple access/collision avoidance (CSMA/CA) mechanism for 802.15.4g to achieve better coexistence with 802.11ah. Furthermore, we present the distributed techniques for 802.15.4g devices to assess 802.11ah interference and switch channel access mode for interference mitigation.

The rest of this paper is organized as follows. Section 2 presents related work. Section 3 evaluates coexistence behavior and issue of 802.11ah and 802.15.4g. We introduce the proposed hybrid CSMA/CA mechanism in Section 4. Section 5 presents distributed methods to estimate 802.11ah interference severity. Section 6 shows network assisted 802.11ah interference severity estimation. In Section 7, we introduce our S1G band coexistence simulator. Performance evaluation of hybrid CSMA/CA mechanism is conducted in Section 8. We conclude our work in Section 9.

2 RELATED WORK

There are existing coexistence technologies developed for conventional 802.15.4 to address its coexistence with 802.11 in 2.4 GHz band. W. Yuan, et al. propose a decentralized approach to mitigate interference by adaptively adjusting ED threshold [8]. E.D.N. Ndih, et al. propose an adaptive backoff mechanism to survive coexistence with 802.11 [9]. X. Zhang, et al. design a cooperative busy tone method via a special device to enable 802.11 to be aware of 802.15.4 transmission [10]. J. Hou, et al. propose a hybrid device to coordinate 802.11 and 802.15.4 transmissions [11]. J.W. Chong, et al. propose an adaptive interference mitigation scheme for 802.15.4 to control its frame length based on the measured 802.11 interference via a hybrid device [12].

Before the work in [6] and [7], to the best of our knowledge, no other existing work addresses the coexistence of 802.11ah and 802.15.4 in the S1G band. The related studies are done either for 802.11ah or 802.15.4g only. B. B. Olyaei, et al. compare the performance of 802.11ah and conventional 802.15.4 in the S1G band. The results reveal that 802.11ah network achieves higher channel efficiency than 802.15.4 network [13]. R. Ma, et al. investigate the coexistence issues of 802.11b and 802.15.4g in 2.4 GHz band. It shows that 802.11b can significantly interfere with 802.15.4g [14]. However, our investigation shows that the existing studies only reveal one side of the story. Table 1 shows majority of available 802.11 and 802.15.4 performance evaluation and conventional coexistence researches.

3 802.11AH AND 802.15.4G COEXISTENCE BEHAVIOR AND ISSUE

Before conducting coexistence performance evaluation using our S1G band coexistence simulator, we briefly introduce the functional differences between 802.11ah and 802.15.4g, which affect the coexistence behavior of 802.11ah and 802.15.4g.

802.11ah defines OFDM PHY and uses the ED-CCA with a threshold of -75 dBm per MHz for coexistence control with other non-802.11 systems. 802.15.4g specifies MR-FSK, MR-OFDM and MR-O-QPSK PHYs and only addresses coexistence among devices using different 802.15.4g PHYs. 802.15.4g ED threshold is lower than -75 dBm, e.g., its ED threshold is in [-100 dBm, -78 dBm] for FSK PHY.

802.11ah channel width is in the unit of MHz, i.e., 1 MHz/2 MHz/4 MHz/8 MHz/16 MHz. However, 802.15.4g channel width is in the unit of kHz, i.e., 200 kHz/400 kHz/600 kHz/800 kHz/1200 kHz. 802.11ah data rate ranges from 150 kbps to 78 Mbps for even one spatial stream. On the other hand, 802.15.4g data rate ranges from 6.25 kbps to 800 kbps.

802.11ah CSMA/CA and 802.15.4g CSMA/CA are much different. 1) 802.11ah allows immediate channel access. 802.15.4g, however, requires backoff no matter how long channel has been idle. 2) 802.11ah backoff is much faster than 802.15.4g backoff due to much smaller parameters as shown in Table 2, where 802.15.4g parameters are for FSK PHY operating in 920 MHz band. 3) 802.11ah requires backoff suspension, i.e., 802.11ah device must perform clear channel assessment (CCA) in each backoff slot and can decrease backoff counter only if the channel is idle. On the other hand, 802.15.4g has no backoff suspension. 802.15.4g device performs CCA after the backoff procedure completes.

The ED threshold, channel width, data rate and first two CSMA/CA features are in favor of 802.11ah. However, the third CSMA/CA feature is in favor of 802.15.4g. Theoretically, an 802.11ah packet can be infinitely delayed, but an 802.15.4g packet has bounded delay.

Based on forementioned functional differences, the purpose of 802.11ah and 802.15.4g coexistence simulation is to explore how network traffic and network size affect the

Table 2: 802.11ah and 802.15.4g CSMA/CA Parameters

802.11ah Param.	Value	802.15.4g Param.	Value
CCA Time	40 μs	phyCCADuration	160 μs
Slot Time	52 μs	UnitBackoffPeriod	1160 μs
SIFS Time	160 μs	AIFS Time	1000 μs
DIFS Time	264 μs	SIFS Time	1000 μs

coexistence behavior of 802.11ah and 802.15.4g as well as what are the critical coexistence issues to be addressed.

We use packet delivery rate and packet latency as metrics to evaluate the coexistence performance. The packet delivery rate is measured as the ratio of number of packets successfully delivered and total number of packets transmitted. The packet latency is measured as time difference from the time at packet transmission process starts to the time at the packet receiving is successfully confirmed. In other words, the packet latency is given by $BackoffTime + DataTXTime + AckWaitingTime + AckRXTime$. The simulation setup is described in section 7.

In Figs. 1 and 2, solid lines represent 802.11ah network performance and dash lines illustrate 802.15.4g network performance. In addition, 50-20-20 indicates 50 nodes for each network, 20 kbps offered load for 802.11ah network, 20 kbps offered load for 802.15.4g network, and so on.

Fig. 1 shows packet delivery rate of 802.11ah network and 802.15.4g network. We have following findings: 1) For all scenarios, 802.11ah network delivers nearly 100% of the packet, which indicates that network traffic and network size have less impact on 802.11ah packet delivery rate. 2) 802.11ah network traffic has impact on 802.15.4g packet delivery rate. 802.15.4g network packet delivery rate decreases as 802.11ah network traffic increases. 3) 802.15.4g network traffic affects more on its packet delivery rate. 802.15.4g network packet delivery rate decreases significantly as its network traffic doubles. 4) The network size has little effect on 802.15.4g network packet delivery rate.

Fig. 2 depicts the corresponding packet latency. We have following observations: 1) For all scenarios, 802.15.4g network achieves similar packet latency, which indicates that 802.15.4g packet is either delivered with the bounded delay or dropped and therefore, network traffic and network size have little impact on 802.15.4g packet latency. 2) 802.11ah network traffic has impact on its packet latency. 802.11ah packet latency increases as its network traffic increases. 3) 802.15.4g network traffic has more impact on 802.11ah packet latency. 802.11ah network packet latency increases more as 802.15.4g network traffic doubles. 4) Network size has major influence on 802.11ah packet latency. 802.11ah packet latency increases significantly as the number of nodes doubles, which verifies that 802.11ah packet can be infinitely delayed.

These results show that 802.11ah network and 802.15.4g network interfere with each other. This observation is different from that drawn by existing studies that only reveal the 802.11ah interference on 802.15.4g. Based on these findings, coexistence technologies need to improve 802.15.4g delivery rate and reduce 802.11ah packet latency.

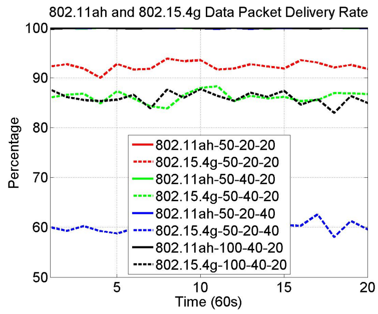


Figure 1: Packet Delivery Rate

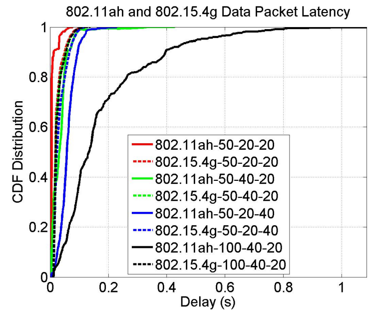


Figure 2: Packet Latency

4 HYBRID CSMA/CA FOR 802.15.4G TO COEXIST BETTER WITH 802.11AH

This section presents the proposed hybrid CSMA/CA for 802.15.4g to improve 802.15.4g delivery rate and reduce 802.11ah packet latency. The proposed hybrid CSMA/CA for 802.15.4g allows 802.14.g device to perform immediate channel access.

An 802.15.4g device cannot communicate with an 802.11ah device. Therefore, 802.15.4g devices cannot coordinate with 802.11ah devices for interference mitigation without special assistance. However, 802.15.4g devices can explore the weakness of 802.11ah devices to increase their channel access opportunity when they detect severe interference from 802.11ah devices. An 802.11ah device must perform backoff process after the busy channel. Before the backoff process, 802.11ah device must wait for a DCF inter frame space (DIFS) (264 μs) time period. This 264 μs waiting time plus random backoff time gives 802.15.4g devices opportunity to start transmission before 802.11ah devices if 802.15.4g devices are allowed to have immediate channel access capability, which is not allowed in the 802.15.4g standard.

To compete with more aggressive 802.11ah for channel access, we propose a hybrid CSMA/CA mechanism for 802.15.4g. Depending on severity of 802.11ah interference, the hybrid CSMA/CA switches between two modes: immediate channel access disabled mode when 802.11ah interference is not severe and immediate channel access enabled mode when 802.11ah interference is severe. In the first mode, the standard 802.15.4g CSMA/CA is applied. In the second mode, the proposed immediate channel access enabled CSMA/CA is employed.

Fig. 3 shows the hybrid CSMA/CA mechanism for

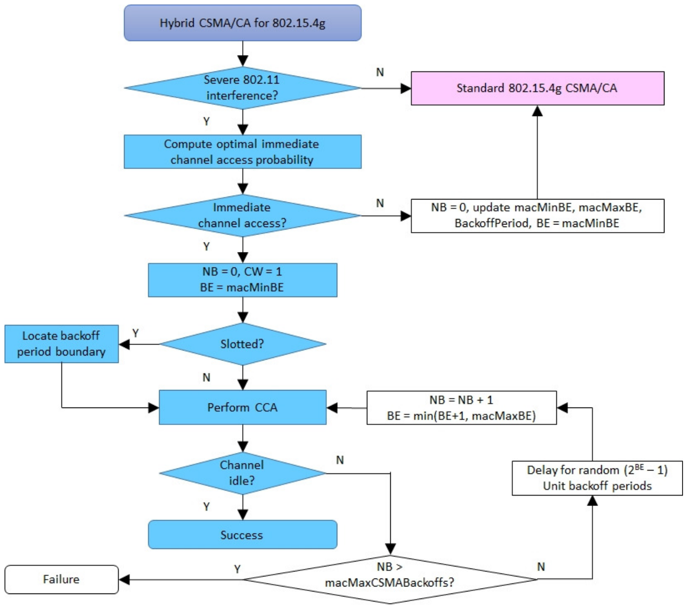


Figure 3: Hybrid CSMA/CA for IEEE 802.15.4g

802.15.4g. To decide a CSMA/CA mode, the hybrid CSMA/CA first determines the severity of 802.11ah interference. If the 802.11ah interference is not severe, the standard 802.15.4g CSMA/CA is applied. If the 802.11ah interference is severe, the immediate channel access enabled CSMA/CA is used. In this mode, the hybrid CSMA/CA enables 802.15.4g devices to have immediate channel access capability. The blue blocks show the flow chart of the immediate channel access. Considering that the immediate channel access by multiple 802.15.4g devices within a neighborhood may also cause collision, the hybrid CSMA/CA computes an optimal probability for stochastic decision making, i.e., perform immediate channel access or backoff.

To compute the optimal probability, an 802.15.4g device first determines number of 802.15.4g neighbors by monitoring neighbor's packet transmission. Assume there are N_g 802.15.4g devices in a neighborhood and each device has probability p to take immediate channel access and probability $1-p$ to perform backoff. Let X denote binomial random variable $\sum_{i=1}^{N_g} X_i^g$, where $X_i^g(i = 1, 2, ..., N_g)$ is random variable representing decision of 802.15.4g neighbor i. Then $P(X = k) = \binom{N_g}{k} p^k (1-p)^{N_g-k}$ and $\mathbb{E}[X] = N_g p$. To avoid collision among 802.15.4g transmissions due to immediate channel access, optimal strategy is that the expected number of 802.15.4g devices that take immediate channel access is one and rest of 802.15.4g devices perform backoff, i.e., $\mathbb{E}[X] = 1$, which gives optimal probability $p_o = \frac{1}{N_g}$.

Based on the optimal probability p_o, the hybrid CSMA/CA decides if immediate channel access or backoff is performed. The **Yes** decision leads to CCA operation. If the CCA returns idle channel, the immediate channel access takes place. The **No** decision leads to backoff. To do so, 802.15.4g device increases backoff parameters to avoid collision with transmission process of immediate channel access device and also give 802.11ah device opportunity to transmit next and therefore, reduces 802.11ah packet latency.

The core of the hybrid CSMA/CA is to determine 802.11ah interference severity. In Sections 5 and 6, we present pure and network assisted distributed methods to estimate 802.11ah interference severity.

5 DISTRIBUTED 802.11AH INTERFERENCE SEVERITY ESTIMATION

This section presents three pure distributed methods for 802.15.4g to estimate 802.11ah interference severity, in which 802.15.4g devices estimate 802.11ah interference severity distributively without any assistance. An 802.15.4g device should select a method for better performance based on the given performance metric.

A) Channel Access Failure Rate Caused by 802.11ah

IEEE 802.15.4g performs carrier sense before starting transmission to check if channel is available. IEEE 802.15.4g can detect other system is transmitting if received signal over IEEE 802.15.4g ED threshold cannot be decoded, and determine channel access failure caused by other system. In this paper, IEEE 802.11ah is assumed to be other system.

Let N_{caf} be the total number of channel access failure observed by an 802.15.4g device for total N_{tx} transmission attempts. N_{caf} can be decomposed into $N_{caf} = N_{caf}^h + N_{caf}^g$, where N_{caf}^h is the number of channel access failure caused by 802.11ah and N_{caf}^g is the number of channel access failure caused by 802.15.4g. 802.15.4g device is able to compute N_{caf}^g by using carrier sense mechanism. To guarantee packet header sensing, 802.15.4g device may start carrier sense early, e.g., start channel sense before backoff counter reaches to zero. Therefore, channel access failure rate caused by 802.11ah R_{caf}^h can be computed as

$$R_{caf}^h = \frac{N_{caf}^h}{N_{tx}} = \frac{N_{caf} - N_{caf}^g}{N_{tx}} \tag{1}$$

If $N_{caf} = N_{caf}^g$, no interference from 802.11ah, else (like $N_{caf} - N_{caf}^g > 0$, 802.11ah presence. 802.15.4g device switches to the immediate channel access.

B) 802.11ah Channel Occupancy Probability

An 802.15.4g device can estimate the channel busy time T_b by continuously sensing channel for a time period T. Its transmission time and reception time are considered as busy time. Its turnaround time is considered as idle time. In addition, 802.15.4g device is able to determine the busy time consumed by 802.15.4g transmissions T_b^g via carrier sense. Therefore, 802.11ah channel occupancy probability P_{tx}^h can be estimated as

$$P_{tx}^h = \frac{T_b - T_b^g}{T}. \tag{2}$$

If P_{ed}^h is higher than 802.15.4g system predetermined threshold, 802.15.4g device switches to the immediate channel access.

C) 802.11ah Energy Detection Ratio

Using energy detection mechanism, an 802.15.4g device can detect energy that is higher than or equal to 802.15.4g ED threshold. Let ED_{total} be the total number of such detection by an 802.15.4g device within a time period T. Using carrier sensing mechanisms, an 802.15.4g device can determine if the detected signal is 802.15.4g signal or not. For 802.15.4g and 802.11ah coexistence, if the signal is not 802.15.4g signal, it is either 802.11ah signal or collided signal. Let ED_{ah} be the number of non 802.15.4g signal detected. 802.11ah energy detection ratio R^h_{ed} can be estimated as

$$R^h_{ed} = \frac{ED_{ah}}{ED_{total}}. \tag{3}$$

If 802.15.4g device detects ED_{ah} during observation time period T and R^h_{ed} is higher than 802.15.4g system predetermined threshold, 802.15.4g device switches to immediate channel access.

6 NETWORK ASSISTED 802.11AH INTERFERENCE SEVERITY ESTIMATION

Some of performance metrics such as packet delivery rate can not be estimated locally by an 802.15.4g device alone and therefore, network assistance is needed. The advantage is that metric can be more accurately assessed. The disadvantage is that network assistance may not be available. Due to the fact that 802.15.4g devices can not distinguish between collision caused by 802.11ah or 802.15.4g, we use the probability of the 802.11ah transmission colliding with 802.15.4g transmission as a metric to estimate 802.11ah interference severity.

This section presents a network assisted distributed method for 802.15.4g devices to estimate 802.11ah interference severity. Using this method, 802.11ah network provides its node distribution, traffic pattern, and the random backoff period length of its nodes to 802.15.4g network. Node distribution and traffic pattern only need to be provided once, and the length of random backoff period needs to be provided repeatedly, since the traffic pattern and terminal location are assumed not to change dynamically for sensor networks that enable IoT application. Node distribution is used to determine 802.11ah neighbors of an 802.15.4g node. The determined number of 802.11ah neighbors, traffic pattern and the length of random backoff period are then used to estimate the collision probability caused by 802.11ah.

An 802.11ah transmission can interfere with an 802.15.4g transmission only if their transmission time periods overlap. Fig. 4(A) shows the collision scenarios.

In the IEEE 802 standards, a data transmission is successful only if its transmission process completes. Therefore, we consider interference effect of 802.11ah transmission process on 802.15.4g transmission process.

An 802.11ah transmission process can interfere with a given 802.15.4g transmission only if the corresponding 802.11ah data arrives within a potential time period. The length of this time period is used to estimate the collision probability caused by 802.11ah.

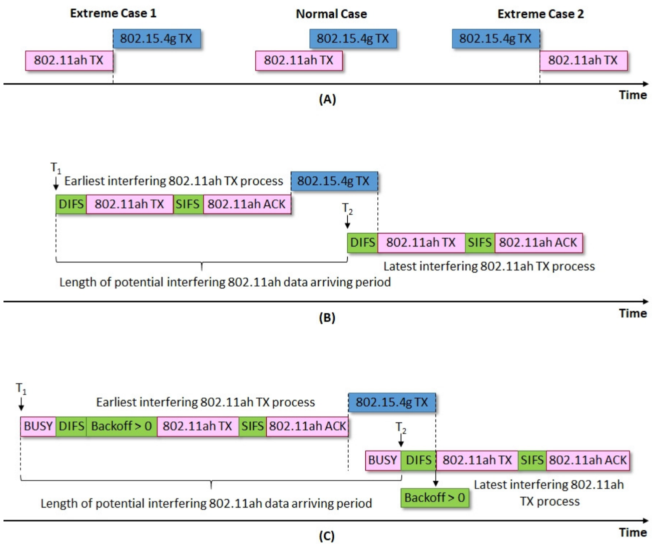


Figure 4: 802.11ah TX interfering with 802.15.4g TX

In this section, we first present a method to estimate the length of potential 802.11ah data arriving time period. Based on 802.11ah traffic pattern and the estimated length of potential 802.11ah data arriving time period, we then compute the probability of 802.11ah transmission process interfering with 802.15.4g transmission process.

A) Potential 802.11ah Data Arriving Time Period

802.11ah channel access can be divided into

1. Immediate access, in which if data arrives, channel is idle and idle channel continues for more than DIFS time period, the data is transmitted without backoff.

2. Deferred access, in which if data arrives, channel is busy, then backoff process is invoked and data transmission is deferred.

An 802.11ah device ignores 802.15.4g transmission if the detected energy level is below 802.11ah CCA-ED threshold, which is -75 dBm for 1 MHz channel, and detects 802.15.4g transmission if the detected energy level is above 802.11ah CCA-ED threshold. Therefore, the 802.11ah interference scenarios can be classified into following four cases:

- **Case-1**: 802.11ah performs immediate channel access and ignores 802.15.4g transmission
- **Case-2**: 802.11ah performs delayed channel access and ignores 802.15.4g transmission
- **Case-3**: 802.11ah performs immediate channel access and detects 802.15.4g transmission
- **Case-4**: 802.11ah performs delayed channel access and detects 802.15.4g transmission

Let T_{gd}, T_{ga}, T_{hd} and T_{ha} be 802.15.4g data transmission time, 802.15.4g ACK transmission time, 802.11ah data transmission time and 802.11ah ACK transmission time, respectively.

For **Case-1**, Fig. 4(B) illustrates the length of potential 802.11ah data arriving period that can cause 802.11ah transmission process interfering with the given 802.15.4g transmission. The period length is given by $T_{im}^{ig} = T_2 - T_1 = T_{hd} + SIFS + T_{ha} + T_{gd}$. It is obvious that the latest interfering 802.11ah transmission process can take place since 802.11ah device ignores 802.15.4g transmission. Is it possible for the earliest interfering 802.11ah transmission process to occur without being detected by 802.15.4g device? Yes, 802.15.4g turnaround time is 1000 μs. 802.11ah SIFS is 160 μs. There are 840 μs left for 802.11ah data transmission and ACK transmission. Even with 1 MHz channel, 802.11ah PHY rate ranges from 300 kbps to 16 Mbps. Using 3 Mbps PHY rate, a 100 byte packet only takes 267 μs. The remaining 573 μs is long enough to transmit 802.11ah ACK.

For **Case-2**, when 802.11ah data arrives for transmission, 802.11ah CCA returns channel status as busy. Therefore, 802.11ah device has to do backoff. Fig. 4(C) depicts the length of potential 802.11ah data arriving period that can cause 802.11ah transmission process interfering with the 802.15.4g transmission. In this case, the earliest interfering 802.11ah transmission process performs backoff with backoff period length greater than zero. The latest interfering 802.11ah transmission process happens to select a zero backoff period length. The period length is given by $T_{df}^{ig} = T_2 - T_1 = \max\{T_{hd}, T_{gd}\} + T_{bo}^h + T_{hd} + SIFS + T_{ha} + T_{gd}$, where $\max\{T_{hd}, T_{gd}\}$ indicates that the busy channel status returned by 802.11ah CCA can be caused by 802.11ah transmission and/or 802.15.4g transmission and T_{bo}^h is the length of random backoff period of 802.11ah device. T_{bo}^h is random with a lower bound 0. For light traffic without backoff suspension, T_{bo}^h has an upper bound $CW_{min} * 52\mu s$, where CW_{min} is typically 15. However, for heavy traffic, T_{bo}^h could be theoretically unbounded.

Combining **Case-1** and **Case-2**, if 802.11ah device ignores 802.15.4g data transmission, the length of potential 802.11ah data arriving time period that can cause 802.11ah transmission process interfering with 802.15.4g data transmission can be estimated as

$$\begin{aligned} T_{itd}^{ig} &= P_i T_{im}^{ig} + (1 - P_i) T_{df}^{ig} \\ &= T_{hd} + SIFS + T_{ha} + T_{gd} \\ &\quad + (1 - P_i)(\max\{T_{hd}, T_{gd}\} + T_{bo}^h), \end{aligned} \tag{4}$$

where P_i is the channel idle probability and can be estimated as $P_i = \frac{T - T_b}{T}$ by using similar approach used in Section 5.

Case-3 is similar as **Case-1**, but in this case, the latest interfering 802.11ah transmission process can not start at the end of 802.15.4g transmission since during 802.15.4g transmission, channel is considered as busy. Therefore, the latest interfering 802.11ah transmission process can only start at the start of 802.15.4g transmission. As a result, the length of potential interfering 802.11ah data arriving time period is $T_{im}^{dt} = T_{hd} + SIFS + T_{ha}$.

Similarly, for **Case-4**, the length of potential interfering 802.11ah data arriving time period is given by $T_{df}^{dt} = \max\{T_{hd}, T_{gd}\} + T_{bo}^h + T_{hd} + SIFS + T_{ha}$.

Combining **Case-3** and **Case-4**, if 802.11ah device detects 802.15.4g data transmission, the length of potential 802.11ah data arriving time period that can cause 802.11ah transmission process interfering with 802.15.4g data transmission can be estimated as

$$\begin{aligned} T_{itd}^{dt} &= P_i T_{im}^{dt} + (1 - P_i) T_{df}^{dt} \\ &= T_{hd} + SIFS + T_{ha} \\ &\quad + (1 - P_i)(\max\{T_{hd}, T_{gd}\} + T_{bo}^h). \end{aligned} \tag{5}$$

B) Collision Probability Caused by 802.11ah

The probability of 802.11ah transmission colliding with 802.15.4g transmission depends on traffic pattern of 802.11ah network. This section considers Poisson data arrival and uniform data arrival traffic scenarios.

In addition, we also assume that 802.15.4g transmission device has N_h 802.11ah neighbors with $N_h > 0$. Otherwise, the 802.15.4g transmission device does not switch CSMA/CA mode and always applies standard CSMA/CA.

B.1) Poisson Data Arrival

Assume 802.11ah device has Poisson data arriving distribution with mean arriving rate λ. In a time period T, the probability an 802.11ah neighbor having no data arriving is $e^{-\lambda T}$ and the probability all 802.11ah neighbors having no data arriving is $e^{-N_h \lambda T}$. Let P_{pd}^h be the probability at least one 802.11ah neighbor having data arriving in time period T, and P_{pd}^h is given by

$$P_{pd}^h = 1 - e^{-N_h \lambda T}. \tag{6}$$

For the immediate access, plugging T_{itd}^{ig} into (6), we obtain the probability 802.11ah transmission process interfering with the given 802.15.4g data transmission as

$$P_{pd}^{ig} = 1 - e^{-\lambda N_h T_{itd}^{ig}}. \tag{7}$$

Similarly, for the deferred access, the probability 802.11ah transmission process colliding with the given 802.15.4g data transmission is given by

$$P_{pd}^{dt} = 1 - e^{-\lambda N_h T_{itd}^{dt}}. \tag{8}$$

Notice that $P_{pd}^{dt} < P_{pd}^{ig}$ since $T_{itd}^{dt} < T_{itd}^{ig}$, which is reasonable because if 802.11ah detects 802.15.4g transmission, it takes action to avoid interference.

Besides interfering with 802.15.4g data transmission, 802.11ah transmission can also interfere with 802.15.4g ACK transmission. 802.15.4g ACK transmission waiting time AIFS is 1000 μs, which is much longer than 802.11ah DIFS time of 264 μs. Therefore, 802.11ah devices can start transmission process in between 802.15.4g data and 802.15.4g ACK. The transmission process can interfere with 802.15.4g ACK transmission.

Consider that 802.15.4g ACK is transmitted only if 802.15.4g data transmission is successful, the probability of 802.15.4g ACK transmission is $1 - P_c^g$, where P_c^g is the 802.15.4g collision probability caused by both 802.11ah transmission and 802.15.4g transmission. 802.15.4g device can compute P_c^g using number of transmission attempts and number of ACK received.

The probability of 802.11ah transmission process interfering with 802.15.4g ACK transmission can be similarly computed as for the 802.15.4g data transmission. In this case, however, the busy channel is caused by 802.15.4g data transmission.

For the immediate access, the probability 802.11ah transmission process colliding with 802.15.4g ACK transmission is given by

$$P_{pa}^{ig} = (1 - P_c^g)(1 - e^{-\lambda N_h T_{ita}^{ig}}), \tag{9}$$

where $T_{ita}^{ig} = T_{hd} + SIFS + T_{ha} + T_{ga} + (1 - P_i)(T_{gd} + T_{bo}^h)$.

For the deferred access, the probability 802.11ah transmission process interfering with 802.15.4g ACK transmission is given by

$$P_{pa}^{dt} = (1 - P_c^g)(1 - e^{-\lambda N_h T_{ita}^{dt}}), \tag{10}$$

where $T_{ita}^{dt} = T_{hd} + SIFS + T_{ha} + (1 - P_i)(T_{gd} + T_{bo}^h)$.

We can also see that $P_{ca}^{dt} < P_{ca}^{ig}$ since $T_{ita}^{dt} < T_{ita}^{ig}$.

Finally, combining all cases, the probability of 802.11ah transmission process colliding with the given 802.15.4g transmission process for Poisson 802.11ah data arriving P_c^p is given by

$$P_c^p = \begin{cases} P_{pd}^{ig} + P_{pa}^{ig}, & \text{if 15.4g data \& ACK ignored} \\ P_{pd}^{dt} + P_{pa}^{ig}, & \text{if only 15.4g data detected} \\ P_{pd}^{ig} + P_{pa}^{dt}, & \text{if only 15.4g ACK detected} \\ P_{pd}^{dt} + P_{pa}^{dt}, & \text{if 15.4g data \& ACK detected.} \end{cases} \tag{11}$$

B.2) Uniform Data Arrival

Assume 802.11ah device has uniform data arriving with time interval T_i^h, i.e., one data arriving per T_i^h time period. For a time period T, the probability of no 802.11ah packet arriving is 0 if $T \geq T_i^h$ and is $((T_i^h - T)/T_i^h)^{N_h}$ if $T < T_i^h$. Therefore, the probability at least one 802.11ah neighbor having data arriving P_{ud}^h is given by

$$P_{ud}^h = \begin{cases} 1, & \text{if } T \geq T_i^h \\ 1 - (\frac{T_i^h - T}{T_i^h})^{N_h}, & \text{if } T < T_i^h. \end{cases} \tag{12}$$

For uniform data arriving, if $T \geq T_i^h$, the hybrid CSMA/CA is always on the immediate access enabled mode. Otherwise, the hybrid CSMA/CA switches mode based on the collision probability caused by 802.11ah.

For the case $T < T_i^h$, P_{ud}^h is simplified as

$$P_{ud}^h = 1 - (\frac{T_i^h - T}{T_i^h})^{N_h}. \tag{13}$$

Similarly as for Poisson data arriving scenario, plugging T_{itd}^{ig}, T_{itd}^{dt}, T_{ita}^{ig} or T_{ita}^{dt} into Eq.(13), we can obtain the probability P_{ud}^{ig}, P_{ud}^{dt}, P_{ua}^{ig} or P_{ua}^{dt}. The probability of the 802.11ah transmission process colliding with the given 802.15.4g transmission process P_c^u is given by

$$P_c^u = \begin{cases} P_{ud}^{ig} + P_{ua}^{ig}, & \text{if 15.4g data \& ACK ignored} \\ P_{ud}^{dt} + P_{ua}^{ig}, & \text{if only 15.4g data detected} \\ P_{ud}^{ig} + P_{ua}^{dt}, & \text{if only 15.4g ACK detected} \\ P_{ud}^{dt} + P_{ua}^{dt}, & \text{if 15.4g data \& ACK detected.} \end{cases} \tag{14}$$

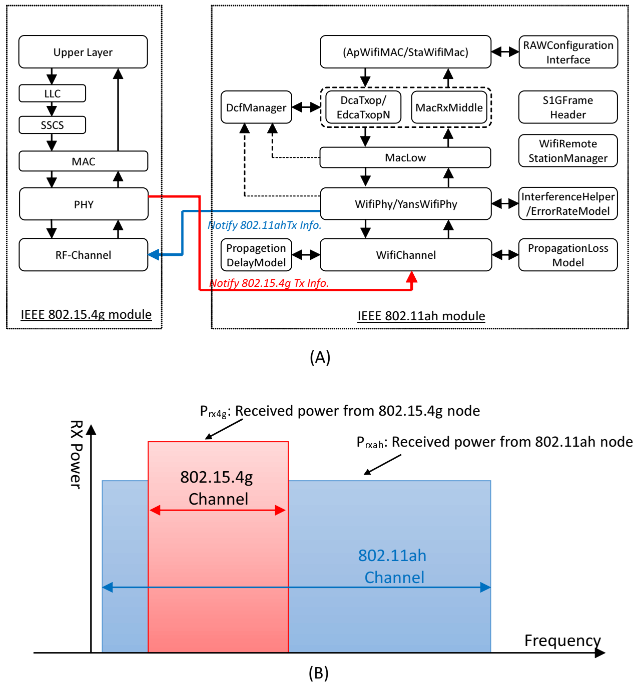


Figure 5: Sub-1 GHz Band Coexistence Simulator Model Interface

7 802.11AH AND 802.15.4G COEXISTENCE SIMULATOR

The existing simulation tools for 802.11 and 802.15.4, e.g., NS-3 [15], MATLAB, QualNet and OMNeT++, do not implement 802.11ah and 802.15.4g. Accordingly, we have developed an NS-3 based coexistence simulator for 802.11ah and 802.15.4g, in which we adopt the third party 802.11ah module [16] and implement 802.15.4g FSK PHY in 920 MHz band. NS-3 (version 3.23) is used because of supported version in [16]. The challenges include the interfacing independent 802.11ah module and 802.15.4g module and the received power conversion.

Fig. 5(A) shows the developed interface between 802.11ah module and 802.15.4g module, where two modules notify each other with their transmission via a *TX Info* (Transmission Information) message that contains device position, transmission duration, transmission power, frequency, bandwidth, antenna gain, etc. Upon receiving *TX Info* message from other party, 802.11ah device and 802.15.4g device first compute the corresponding RX power P_{rx4g} and P_{rxah}, respectively, as shown in Fig. 5(B), where same transmission power is assumed. In other words, 802.11ah device computes 802.15.4g received power P_{rx4g} as if it was an 802.15.4g device and 802.15.4g device computes

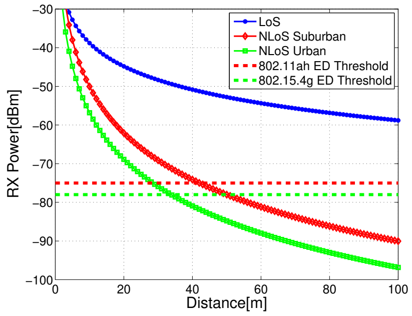


Figure 6: ITU-R P.1411-9 Propagation Model

802.11ah received power P_{rxah} as if it was an 802.11ah device. Using the received power computed, 802.11ah device and 802.15.4g device compute interference power level from other party as

$$\begin{aligned} P_{int}^{4g} &= P_{rxah}[dBm] - 10\log_{10}(CH_{ah}/CH_{4g})[dBm], \\ P_{int}^{ah} &= P_{rx4g}[dBm], \end{aligned} \tag{15}$$

where P_{int}^{ah} is interference power to 802.11ah from 802.15.4g transmission, P_{int}^{4g} is interference power to 802.15.4g from 802.11ah transmission, CH_{ah} and CH_{4g} represent the channel width of 802.11ah channel and 802.15.4g channel, respectively. Using the interference power level and transmission duration, 802.11ah device and 802.15.4g device perform the enhanced CCA operation such that if the interference power is above the corresponding CCA-ED threshold, the channel status is considered as busy no matter what channel status is returned by their respective CCA operation.

Propagation model is another key component for practical simulation. NS-3 implements eight propagation models designed for general use scenarios without considering the emerging IoT applications. Both 802.11ah and 802.15.4g target the outdoor applications such as smart utility and smart city. Therefore, we adopt ITU-R P.1411-9 model for propagation between terminals located from below roof-top height to near street level. The median value of the Non-Line-of-Sight (NLoS) loss is given by

$$L_{NLoS}^{median}(d) = 9.5 + 45\log_{10} f + 40\log_{10}(d/1000) + L_{urban}, \tag{16}$$

where f is the frequency, L_{urban} depends on the urban category and is 0 dB for suburban, 6.8 dB for urban, and d is the distance. Fig. 6 shows the propagation loss of LoS model, Suburban NLoS model and Urban NLoS model for transmission power of 13 dBm. With -78 dBm ED threshold, the intersection of the red curve and green dash line represents the effective energy detection distance for 802.15.4g, which is about 50 meters for Suburban NLoS model and 34 meters for Urban NLoS model. For 802.11ah with -75 dBm ED threshold, the corresponding distances are 42 meters and 28 meters, respectively.

Table 3: Simulation Scenarios

	Network Size [node]		Offered Load [kbps]		Propagation Model
	11ah	15.4g	11ah	15.4g	
Scenario-1	50	50	20	20	Suburban NLoS
Scenario-2	50	50	40	20	
Scenario-3	50	50	20	40	
Scenario-4	100	100	40	20	
Scenario-5	100	100	40	20	Urban NLoS

8 HYBRID CSMA/CA PERFORMANCE EVALUATION

In this section, we evaluate the performance of Hybrid CSMA/CA proposed in Section 4 to compare with standard 802.15.4g CSMA/CA. We adopt the simulation parameters recommended by IEEE 802.19 Working Group [17]. Table 4 shows simulation parameters for IEEE 802.11ah and IEEE 802.15.4g coexistence performance. The frequency is in 920 MHz band, transmission power is 13 dBm, 1 MHz channel for 802.11ah, 400 kHz channel for 802.15.4g, 802.11ah OFDM PHY rate is 300 kbps and 802.15.4g FSK PHY rate is 100 kbps. And, ITU-R P.1411-9 propagation model is employed in the simulations. We use 802.11ah energy detection ratio method to assess 802.11ah interference severity. We define PDR (Packet Delivery Rate) and packet latency as metrics to evaluate the coexistence performance.

Five typical scenarios for 50-node and 100-node are simulated with simulation conditions in [17]. Table 3 shows the simulation cases. One 802.15.4g network consists of 50 or 100 nodes uniformly deployed in a circle centered at PANC (Personal Area Network Coordinator) with radius of effective energy detection distance. The PANC is located at (0, 0). Three 802.11ah networks are deployed inside 802.15.4g network with each 802.11ah network having 17 or 33 nodes uniformly distributed in a circle centered at corresponding AP with radius of effective energy detection distance. Based on propagation model, three APs are located at (8, 0), (-4, 6.928), (-4, -6.928) and (6, 0), (-3, 5.196), (-3, -5.196), respectively. The offered network load is 20 kbps or 40 kbps. The offered network load is uniformly distributed among network nodes. The packet size is 100 bytes.

In addition to five typical scenarios, individual performance of IEEE 802.11ah and IEEE 802.15.4g was conducted for $(node, offeredload) = (50\,nodes, 20\,kbps)$. Packet arrival rate in a single network for is 100% for IEEE 802.11ah and 98.5 % for IEEE 802.15.4g, respectively. Packet latency at CDF 0.9 is 10 ms for IEEE 802.11ah and 40 ms for IEEE 802.15.4g.

Scenario-1: The offered load for both networks is 20 kbps, i.e., 400 bps offered load per node, which leads to 0.13 % duty cycle for 802.11ah node and 0.4 % duty cycle for 802.15.4g node. These duty cycles are much lower than the 10 % duty cycle specified in ARIB STD T108 standard [3]. With 100 bytes of packet size, each node generates 0.5 packet per second. For both standard CSMA/CA and hybrid CSMA/CA, Fig. 7 shows that 802.11ah network delivers 100 % of the packet. The standard CSMA/CA delivers 92.37 % of 802.15.4g packet. The hybrid CSMA/CA delivers 95.77 % of

Table 4: Simulation parameters for IEEE 802.11ah and IEEE 802.15.4g coexistence performance

Parameters	Value [Unit]	Note
Network offered load	20-40 kbps	11ah
Network offered load	20-40 kbps	15.4g
Tx Power	20 mw	11ah & 15.4g
11ah Bandwidth	1 MHz	11ah
15.4g Bandwidth	400 kHz	15.4g
aSlotTime	52 usec	11ah
aSIFSTime	160 usec	11ah
aCCATime	<40 usec	11ah
aRxTxTurnaroundTIme	Less than 5 usec	11ah
CW (min, max)	15, 1023	11ah
phyCCADuration	140 usec	15.4g
aTurnaroundTime	1000 usec	15.4g
Rx to Tx TrunaroundTime	300 usec or more, 1000 usec or less	15.4g
Tx to Rx TurnaroundTIme	Less than 300 usec	15.4g
macMinLIFSPeriod	1000 usec	15.4g
aUnitBackoffPeriod	1140 usec	15.4g
macAckWaitDuration	5 ms	15.4g
macMaxBE	3 to 8 (Default 5)	15.4g
macMinBE	0 to macMaxBE (Default 3)	15.4g
macMaxCSMABackoffs	0 to 5 (Default 4)	15.4g
macMaxFrameRetries	0 to 7 (Default 4)	15.4g

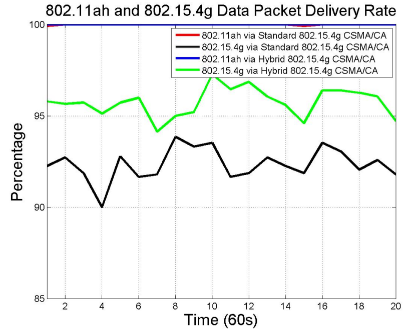


Figure 7: Scenario-1: Packet Delivery Rate

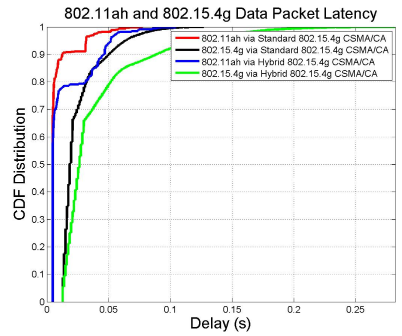


Figure 8: Scenario-1: Packet Latency

802.15.4g packet, i.e., 3.4 % improvement without degrading 802.11ah packet delivery. Fig. 8 shows that packet latency for both 802.11ah and 802.15.4g, Standard CSMA/CA achieves shorter packet latency than the hybrid CSMA/CA due to less 802.15.4g packet delivered. 802.11ah has shorter packet latency than 802.15.4g. In this case, the hybrid CSMA/CA increases 802.11ah packet latency slightly.

Scenario-2: The offered load is 40 kbps for 802.11ah network and 20 kbps for 802.15.4g network, i.e., the offered load is 800 bps for 802.11ah node and 400 bps for 802.15.4g node, which leads to 0.26 % duty cycle and 0.4 % duty cycle, respectively. These duty cycles are much lower than the 10 % duty cycle limit. Each 802.11ah node generates 1 packet per second and each 802.15.4g node generates 0.5 packet per second. Fig. 9 shows that both standard CSMA/CA and hybrid CSMA/CA deliver near 100 % of 802.11ah packet. The hybrid CSMA/CA improves 802.15.4g packet delivery rate from 86.2 % given by standard CSMA/CA to 90.7 %. This 4.5 % improvement is done without degrading 802.11ah packet delivery. It indicates that as 802.11ah network traffic increases, the hybrid CSMA/CA provides more improvement on 802.15.4g packet delivery rate. Fig. 10 shows that 802.11ah and 802.15.4g have similar packet latency. For 802.15.4g, standard CSMA/CA achieves slightly shorter packet latency than the hybrid CSMA/CA due to less 802.15.4g packet delivered. However, the hybrid CSMA/CA maintain overall 802.11ah packet latency, i.e., the hybrid CSMA/CA does not degrade 802.11ah performance.

Scenario-3: The offered load is 20 kbps for 802.11ah network and 40 kbps for 802.15.4g network, i.e., the offered load is 400 bps for 802.11ah node and 800 bps for 802.15.4g node, which leads to 0.13 % duty cycle and 0.8 % duty cycle, respectively. These duty cycles are much lower than the 10 % duty cycle limit. Each 802.11ah node generates 0.5 packet per second and each 802.15.4g node generates 1 packet per second. Fig. 11 shows that both standard CSMA/CA and hybrid CSMA/CA deliver near 100 % of 802.11ah packet. The hybrid CSMA/CA improves 802.15.4g packet delivery rate from 59.8 % given by standard CSMA/CA to 61.3 %, i.e., 1.5 % of improvement without degrading 802.11ah packet delivery. It indicates that as 802.15.4g traffic increases, the hybrid CSMA/CA provides less improvement on 802.15.4g packet delivery rate. Fig. 12 shows that 802.11ah delay packet longer than 802.15.4g does due to high 802.15.4g network traffic. For both 802.11ah and 802.15.4g, standard CSMA/CA achieves slightly shorter packet latency than the hybrid CSMA/CA due to less 802.15.4g packet delivered. In this case, the hybrid CSMA/CA delays 802.11ah further longer.

Scenario-4: In this case, each of 802.11ah network and 802.15.4g network has 100 nodes. The offered load, the duty cycle and the number of packet per second are same as in the Scenario-2 of the 50-node scenario. Fig. 13 shows that both standard CSMA/CA and hybrid CSMA/CA deliver near 100 % of 802.11ah packet. For 802.15.4g, the hybrid CSMA/CA improves packet delivery rate from 86.1 % given by standard CSMA/CA to 92.9 % without degrading 802.11ah packet delivery rate. This 6.8 % improvement is better than 4.5 % improvement in the Scenario-2 of 50-Node scenario. Fig. 14 shows that 802.15.4g achieves lower packet latency than 802.11ah. For 802.15.4g, the standard CSMA/CA delays packet shorter than the hybrid CSMA/CA due to less 802.15.4g packet delivered. For 802.11ah, however, the hybrid CSMA/CA achieves shorter packet delay than the standard CSMA/CA does. This is because for the standard CSMA/CA, the range of the 802.15.4g backoff period length

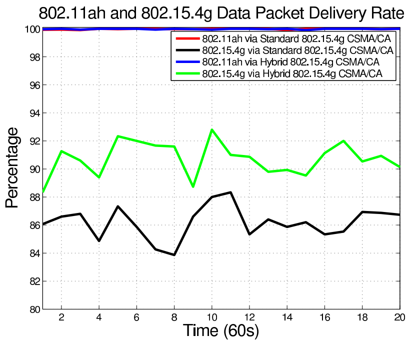


Figure 9: Scenario-2: Packet Delivery Rate

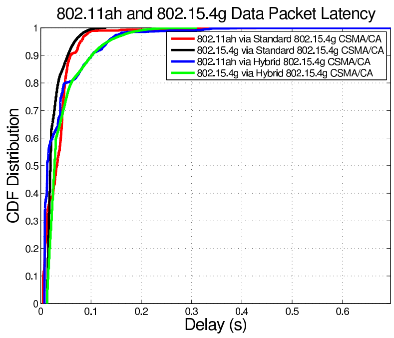


Figure 10: Scenario-2: Packet Latency

is smaller, which results in more concentrated 802.15.4g packet transmission and therefore, causes the longer delay of 802.11ah transmission. On the other hand, the hybrid CSMA/CA allows the longer range of the 802.15.4g backoff period, which spreads 802.15.4g transmission and gives 802.11ah opportunity to transmit early. Therefore, 802.11ah achieves shorter packet delay. In this case, the hybrid CSMA/CA not only improves 802.15.4g packet delivery rate but also improves 802.11ah packet latency. This case also demonstrates that as the number of network node increases, the hybrid CSMA/CA becomes more effective.

Scenario-5: In this case, network size, the offered load, the duty cycle and the number of packet per second are same as in Scenario-4. The difference from Scenario-4 is propagation model from Suburban NLoS model to Urban NLoS Propagation model. Again, Fig 15 shows that both standard CSMA/CA and hybrid CSMA/CA deliver near 100 % of 802.11ah packet. The standard CSMA/CA delivers 76.87 % of 802.15.4g packet. The hybrid CSMA/CA delivers 82.18 % of 802.15.4g packet, i.e., 5.31 % improvement without degrading 802.11ah packet delivery. However, using Urban NLoS model, both standard CSMA/CA and hybrid CSMA/CA achieve lower packet delivery rate compared with the corresponding results using Suburban NLoS model (as on Scenario-4) due to the higher node density, which causes more channel access failure for 802.15.4g nodes. It indicates that as node density increases, 802.15.4g packet delivery rate decreases. Fig. 16 shows that packet latency for both 802.11ah and 802.15.4g, the standard CSMA/CA delays packet less than the hybrid CSMA/CA due to more packet drop by standard CSMA/CA. However, compared with the Suburban NLoS mode case (as on Scenario-4), Urban NLoS model has shorter packet delay due to less 802.15.4g packet delivery.

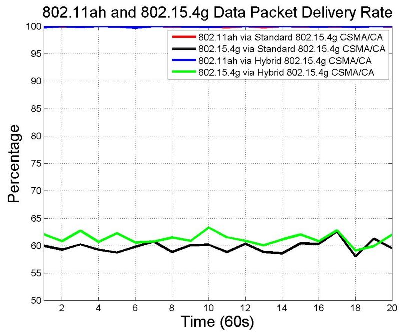


Figure 11: Scenario-3: Packet Delivery Rate

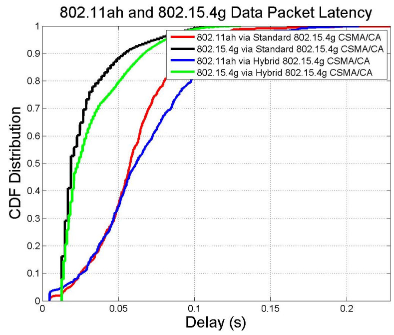


Figure 12: Scenario-3: Packet Latency

Table 5: Packet Delivery Rate Comparison

	802.11ah		802.15.4g		
	Standard	Hybrid	Standard	Hybrid	Diff.
Scenario-1	100 %	100 %	92.4 %	95.8 %	3.4 %
Scenario-2	100 %	100 %	86.2 %	90.7 %	4.5 %
Scenario-3	100 %	100 %	59.8 %	61.3 %	1.5 %
Scenario-4	100 %	100 %	86.1 %	92.9 %	6.8 %
Scenario-5	100 %	100 %	78.8 %	82.1 %	5.3 %

In summary of all network traffic and network size scenarios, simulation results show that the proposed hybrid CSMA/CA improves 802.15.4g packet delivery rate without degrading 802.11ah packet delivery rate. In some case, it can improve performance of both 802.11ah network and 802.15.4g network. As number of the node increases, the hybrid CSMA/CA demonstrated more superiority. In addition, the Suburban NLoS model outperforms Urban NLoS model. Packet delivery rate of 802.11ah and 802.15.4g for both standard 802.15.4g CSMA/CA and the proposed hybrid CSMA/CA for 802.15.4g are shown in Table 5.

9 CONCLUSION

The heterogeneous wireless technologies developed for IoT applications increase the coexistence potential and present coexistence challenges. This paper takes IEEE 802.15.4g and IEEE 802.11ah as target technologies to investigate the Sub-1 GHz band coexistence. We evaluated 802.15.4g and 802.11ah coexistence behavior and identified 802.15.4g packet delivery rate and 802.11ah packet latency

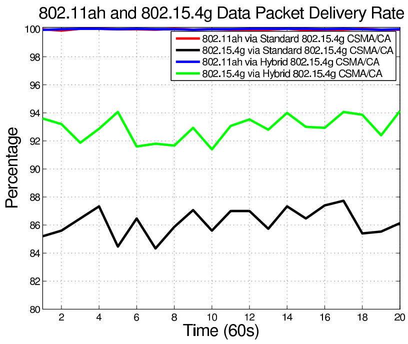


Figure 13: Scenario-4: Packet Delivery Rate

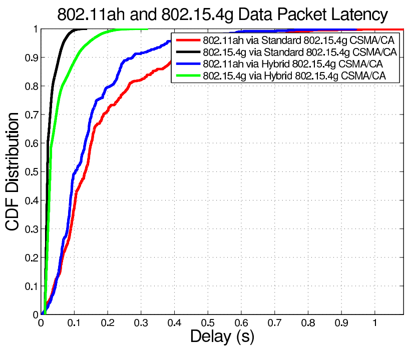


Figure 14: Scenario-4: Packet Latency

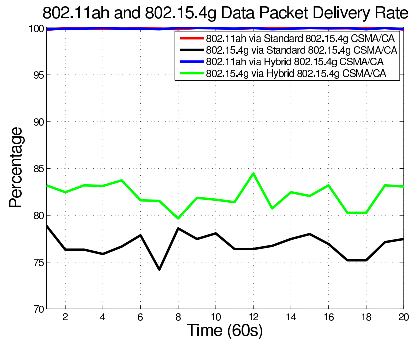


Figure 15: Scenario-5: Packet Delivery Rate

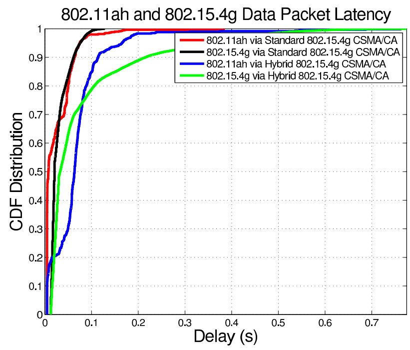


Figure 16: Scenario-5: Packet Latency

as the coexistence issues to be addressed. Accordingly, we proposed a hybrid CSMA/CA mechanism for 802.15.4g to achieve better coexistence with 802.11ah. To contend for channel access with more aggressive 802.11ah, the hybrid CSMA/CA allows 802.15.4g to perform immediate channel access. Two classes of the distributed methods are introduced for 802.15.4g devices to estimate the severity of 802.11ah interference and switch the CSMA/CA mode for interference mitigation. Using the developed Sub-1 GHz band coexistence simulator with use case scenarios and parameters proposed by IEEE 802.19.3 Task Group, we conducted the performance analysis of the proposed hybrid CSMA/CA. Compared with the standard 802.15.4g CSMA/CA, simulation results show that the hybrid CSMA/CA can improve 802.15.4g packet delivery rate by 6.8% without degrading 802.11ah packet delivery rate in our scenario. As the number of nodes in the network traffic increases under the same conditions, the hybrid CSMA/CA can also reduce 802.11ah packet latency.

REFERENCES

[1] IEEE 802.11ah-2016, "IEEE Standard for Information technology–Telecommunications and information exchange between systems - Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Sub 1 GHz License Exempt Operation," IEEE 802.11ah-2016, IEEE, 2016.

[2] IEEE 802.15.4g-2012, "IEEE Standard for Local and metropolitan area networks–Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 3: Physical Layer (PHY) Specifications for Low-Data-Rate, Wireless, Smart Metering Utility Networks," IEEE 802.15.4g-2012, IEEE, (2012).

[3] ARIB STD-T108, "920MHz-Band Telemeter, Telecontrol and Data Transmission Radio Equipment," https://www.arib.or.jp/english/html/overview/doc/5-STD-T108v1_0-E1.pdf, (2012).

[4] B. A. Rolfe, "PAR as approved by REVCOM Dec 2019," IEEE802.19-18-0093r0, https://mentor.ieee.org/802.19/documents/2018/ IEEE 802.19, (2018).

[5] Y. Nagai, J. Guo, T. Sumi, P. Orlik and H. Mineno, "Hybrid CSMA/CA for Sub-1 GHz Frequency Band Coexistence of IEEE 802.11ah and IEEE 802.15.4g," 14th International Workshop on Informatics (IWIN), INFSOC, (2020).

[6] J. Guo and P. Orlik, "Self-transmission control in IoT over heterogeneous wireless networks," 2017 Ninth International Conference on Ubiquitous and Future Networks, pp. 898-903, Milan, (2017).

[7] Y. Liu, J. Guo, P. Orlik, Y. Nagai, K. Watanabe and T. Sumi, "Coexistence of 802.11ah and 802.15.4g networks," 2018 IEEE Wireless Communications and Networking Conference, pp. 1-6, Barcelona, (2018).

[8] W. Yuan, J. M. G. Linnartz and I. G. M. M. Niemegeers, "Adaptive CCA for IEEE 802.15.4 Wireless Sensor Networks to Mitigate Interference," 2010 IEEE Wireless Communication and Networking Conference, Sydney, NSW, pp. 1-5, (2010).

[9] E. D. N. Ndih and S. Cherkaoui, "Adaptive 802.15.4 backoff procedure to survive coexistence with 802.11 in extreme conditions," 2016 13th IEEE Annual Consumer Communications and Networking Conference, Las

Vegas, NV, pp. 556-561, (2016).

[10] X. Zhang and K. G. Shin, "Enabling Coexistence of Heterogeneous Wireless Systems: Case for ZigBee and WiFi," the 12 th ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, (2011).

[11] J. Hou, B. Chang, D.-K. Cho, and M. Gerla, "Minimizing 802.11 Interference on Zigbee Medical Sensors," the Fourth International Conference on Body Area Networks, ICST, (2009).

[12] J. W. Chong, C. H. Cho, H. Y. Hwang, and D. K. Sung, "An Adaptive WLAN Interference Mitigation Scheme for ZigBee Sensor Networks," International Journal of Distributed Sensor Networks, vol. 11, no. 8, pp. 1-16, (2015).

[13] B. B. Olyaei, J. Pirskanen, O. Raeesi, A. Hazmi, and M. Valkama, "Performance Comparison Between Slotted IEEE 802.15.4 and IEEE802.11ah in loT Based Applications," in 1st International Workshop on Internet of Things Communications and Technologies, IEEE, pp. 332-337, (2013).

[14] R. Ma, S. Chen, H. Chen and W. Meng, "Coexistence of Smart Utility Networks and WLANs in Smart Grid Systems," IEEE Transactions on Wireless Communications, vol. 15, no. 12, pp. 8313-8324, (Dec. 2016).

[15] The ns-3 network simulator, http://www.nsnam.org/.

[16] L. Tian, L. Latre, and J. Famaey, "Implementation and validation of an IEEE 802.11ah module for NS-3," in Workshop in NS3, ACM, (2016).

[17] Y. Nagai, J. Guo, T. Sumi, P. Orlik, and H. Mineno, "S1G Coexistence Simulation Profile," IEEE802.19-19-0021r2, https://mentor.ieee.org/802.19/documents/2019/, IEEE 802.19, (May 2019).

(Received October 29, 2020)
(Accepted July 6, 2021)

Yukimasa Nagai received B.E. degree and M.E. degree in Information and Communication Engineering from The University of Electro-Communications, Tokyo, Japan in 1998 and 2000, respectively. He has been working at Information Technology R&D Center of Mitsubishi Electric Corporation (2000 - 2016, 2020-) and Mitsubishi Electric Research Laboratories (2016 - 2020). His research work has covered various topics on wireless communication systems, signal processing and connected car, and he has acted as a key part of getting standardization groups formed in Wi-Fi Alliance and IEEE 802.11/15/19. Wi-Fi Alliance DSRC MTG Vice Chair (2015 - 2018), Wi-Fi Alliance Automotive MSTG Vice Chair (2013 - present), IEEE Senior member, and IPSJ member.

Jianlin Guo is a Senior Principal Research Scientist at Mitsubishi Electric Research Laboratories in Cambridge, Massachusetts, USA. He received his Ph.D. in Applied Mathematics in 1995 from University of Windsor, Windsor, Ontario, Canada. His research interests include coexistence of the heterogeneous wireless networks, time sensitive networking, edge computing for vehicular networks, networked control systems, anomaly detection in factory automation, routing and resource management in wireless IoT networks, and nonlinear stability of convection in porous medium. He is IEEE Senior Member and IEEE 802.19.3 Task Group Technical Editor.

Takenori Sumi received B.E. degree and M.E. degree in Faculty of Engineering of Doshisha University, Kyoto, Japan in 2005 and 2007, respectively. He has been working as a senior researcher at Information Technology R&D Center of Mitsubishi Electric Corporation from 2007. His current research interests include spectrum sharing, wireless coexistence and V2X. He is also engaged in the Wi-Fi architect and network architect for embedded system. IEEE Senior member and IEICE member.

Philip Orlik received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York at Stony Brook. In 1999 he earned his Ph. D. in electrical engineering also from SUNY Stony Brook. Since 2000 he has been with Mitsubishi Electric Research Laboratories Inc. located in Cambridge, MA where he is currently the Manager of the Signal Processing Group. His primary research focus is on advanced wireless and wired communications, sensor/IoT networks. Other research interests include vehicular/car-to-car communications, mobility modeling, performance analysis, and queuing theory.

Hiroshi Mineno received his B.E. and M.E. degrees from Shizuoka University, Japan in 1997 and 1999, respectively. In 2006, he received his Ph.D. degree in information science and electrical engineering from Kyushu University, Japan. Between 1999 and 2002, he was a researcher in the NTT Service Integration Laboratories. In 2002, he joined the Department of Computer Science of Shizuoka University as an Assistant Professor. He is currently a Professor. His research interests include Intelligent IoT systems as well as heterogeneous network convergence. He is a senior member of IEEE, IEICE and IPSJ, a member of ACM and the Informatics Society.

Regular Paper

Sound Radiation of Human Voices in Singing

Orie Abe† and Rolf Bader‡

†Information Communication Media Laboratory, Japan
‡Institute for Systematic Musicology, University of Hamburg, Germany
†orie_deutschland@hotmail.com
‡R_Bader@t-online.de

Abstract - A topic addressed in this paper is in general how the sound energy of human voices radiates in singing. In particular, we were interested in whether sound radiation from singers is restricted to the mouth. Furthermore, it was investigated whether there are differences among vowels, pitches, and between classical and nonclassical styles of singing. For this study, the sound energy of seven trained singers from four musical genres was measured at 15 parts of the upper body while they sang five vowels (a – e – i – o – u) at different pitches in accordance with their voice classification. In this investigation, a microphone array of 121 microphones was used and the gained information was analyzed using the Minimum Energy Method and Mathematica by which the radiating field and radiation patterns could be visualized. The results showed that the main source of the radiation energy of singing voices is the mouth, as expected, but strong energy can arise from other parts of the body as well. Furthermore, the research findings indicated that the radiation energy of singing voices change according to the frequency analyzed, the vowel, the pitch that is sung, and the person who sings.

Keywords: Human Voice, Singing Voice, Sound Radiation, Microphone Array, Minimum Energy Method

1 INTRODUCTION

In our daily life, we receive various acoustical signals through the air, including the sound of the human voice. The human voice is one of the most important communication mediums to convey feelings or impressions of the speaker or singer [12]. Each voice art, such as "speaking voice" or "singing voice," and each voice quality, like pressed or husky, occurs from the body; i.e., the same organs (e.g. vocal cords, vocal tract etc.) are used for each production of the voice. However, there are some differences not only between speaking voice and singing voice, but also among voice qualities which arise due to their desired purpose. For example, singers must produce a specific volume, and therefore they use their body as a "musical instrument," adjusting body parts in a desired way. Especially, classical singers must acquire the ability to surpass the volume level of a loud orchestral accompaniment, for their singing voice to reach the audience. Thus it is essential to acquire expertise for the development of their vocal ability. In contrast, nonclassical singers usually use a microphone while singing, so that the vocal instrument is used for communicative purpose just like in spoken communication, but with music.

However, the fact is that each singer's voice comes from the body while singing and that the mouth is the main source of its radiation energy. In the process of singing, the mouth is working as a mouth piece, comparable to the mouth piece of a trumpet [1]. But the sound can still be audible, even when it is hummed with closed lips. The energy has to be radiated somehow, otherwise the sound would barely be audible. Another interesting fact is that singers sometimes remark that certain parts of their body tend to vibrate while singing. What about such vibrations? Are these vibrations radiated and what is their strength in the total energy of the singing voice? Or is sound radiation from singers actually restricted to the mouth?

In addition, singers also remark that the strength of such vibrations vary in correlation with the vocal technique, the pitch, and the vowel used. Hence it can be expected that the pattern of sound radiation changes if a singer adopts a different vocal technique or sings a different vowel or tone at a different pitch. But, is that actually so, as most musical instruments have complex patterns of sound radiation, which change with direction, pitches played, and other factors?

Therefore, for a better understanding about the origin of radiation energy of the singing voice, the topic addressed in this study is mainly how strong the radiation energy of the singing voice from the upper body is, in comparison to the mouth radiation, and whether the radiation energy of a singer's voice varies according to vocal technique, pitch, and vowel.

2 PREVIOUS RESEARCH METHODS

In the past, the topic whether the vocal sound energy is restricted to the mouth only, has been studied repeatedly (e.g. [2]-[7]). These studies found out that the sound vibration depends on the vowels because of the mouth aperture size and on the pitch. Furthermore, stronger vibration was also observed at loud vocal sound. In addition, there are also many researches on voice directivity of singing voices (a review of recent studies can be found [12]). But in this work, we focus just on the vibrations of the singer's body while singing.

The previous studies on this topic have mostly been executed using an accelerometer, with which the energy of each desired part of the body has been measured. In order to yield an accurate result from each desired part of the body, a measurement using an accelerometer is certainly a suitable research method. But our investigation was executed by means of a microphone array with 121 microphones, the so-called Acoustic Camera, because using a microphone array has three advantages which would be ideal for the purpose of this research [8]:

- It is a non-invasive method.
- It enables us to see the radiation field of the radiating source.
- The recorded data can be used to back-propagate the sound field to the surface of the radiating source, so the sound pressure level from the mouth and the vibrating body structures of the singers are directly comparable.

The method which uses a microphone array is commonly used in musical instrumental acoustics nowadays. However, in vocal acoustics, the use of a microphone grid seems to be standard and the research method using an acoustic camera is still relatively new, particularly in the size with 121 microphones.

3 AIM, METHOD AND RESEARCH MATERIALS

3.1 Aim

Mainly, this study attempted to find an answer to the following questions:

- Is it true that the main sound energy of the singing voice only radiates from the mouth? Or are there other body parts involved in the sound radiation of the singing voice?
- If so, are there recognizable differences depending on vocal music genres /singing techniques and vowels?
- How strong is the sum of the radiated energy from all the measured parts of the body in relation to the mouth radiation?

Additionally, another analysis was executed, because the radiation energy originated from the corners of the mouth has presumably a strong influence on the total energy, due to the proximity of these parts to the mouth. Therefore, the question is:

- What about the radiation energy excluding the corners of the mouth?

3.2 Method

The microphone array, which was employed in our investigation, consists of 121 microphones (11 x 11) and the array spacing is a regular grid with a grid constant of 3.9 cm. This construction enables a symmetric visualization of the radiating field later. The microphones record simultaneously with a sampling frequency of 48 kHz, thus covering whole human hearing range up to 20 kHz. This is a very important factor for measurements of musical instruments including singing voice, because musical instruments often radiate high frequency and initial transients are often the most important part of the sound [8].

After the recording, the data obtained by the array were used to back-propagate the sound field to the radiating source surface by means of a near-field method, the so-called Minimum Energy Method [8], for reconstructing a sound pressure field at a radiating surface of musical instruments including the human body for voice research. It samples the source plane by as many equivalent sources as the microphones present in the Acoustic Camera. This method enables a reconstruction of sound pressure fields and a visualization of an overall radiation directivity of a vibrating geometry.

For the analysis, a code written in Mathematica was applied to all the data and the vibrations were analyzed on a total of 15 parts of the upper body (mouth / chin / throat / left and right clavicles / sternum / nose / nasal bone / left and right corners of the mouth / left and right cheeks / forehead / left and right lower eyelids). This setting made it possible to show energy values of the voice radiation from the singer's upper body, including the phase angles and all statical procedures are based on those visualized results. The data are reconstructed and visualized on a photo of the human upper body from the head to the chest.

Furthermore, the strongest radiated area was color-marked: the radiation is adjusted up to -6 dB and the intensity of the radiation energy is visualized by colors (the brighter the color, the stronger the radiated energy). Usually the radiation from the mouth is the strongest, so that the radiated energies of all single frequencies were therefore normalized to 0 dB at the mouth, but there were some figures where the energy value at the mouth is shown in -1 dB. This means that the revealed energy value is not quite 0 dB, so somewhat stronger than 0 dB.

Both, Acoustic Camera and the Minimum Energy Method were developed at the Institute for Systematic Musicology, University of Hamburg, and all information about these recording and analysis techniques can be found in the publications of the developer Rolf Bader (Detailed description about the Acoustic Camera and the Minimum Energy Method as well as exemplary measurements can be found e.g. in [8]-[14]).

3.3 Research Materials

Seven trained singers — three classical, two musical theater, and two popular singers — participated in this study and one popular singer's voice (subject: VS) was recorded twice by means of popular singing technique and by means of Soul singing technique:

1. CH (classical, female, alto)
2. SE1 (classical, female, soprano)
3. JR (classical, male, bass)
4. SS (musical theater, female, mezzo soprano)
5. TF (musical theater, male, tenor)
6. SE2 (Pop, female, mezzo soprano)

7. VS (Pop, female, mezzo soprano)

8. VS (Soul, female, mezzo soprano)

For the study, after a short warming-up vocal exercise, each subject sang five vowels a/e/i/o/u at the following fundamental frequencies;

- 90 Hz (bass), analyzed up to 1.3 kHz (15 partials)
- 120 Hz (males), analyzed up to 2.5 kHz (21 partials)
- 180 Hz (alto/mezzo), analyzed up to 3 kHz (17 partials)
- 250 Hz (all subjects), analyzed up to 4 kHz (16 partials)
- 380 Hz (tenor), analyzed up to 4.5 kHz (12 partials)
- 500 Hz (females), analyzed up to 5 kHz (10 partials)

Each vowel was recorded separately and max. 2 seconds of phonation. In order to get good data, the vowels were repeatedly measured two or three times in the order of a-e-i-o-u, respectively. The measurement was executed in an anechoic chamber. For the recording, the microphone array was attached to the front of a stand and adjusted for the height of the singer, so that their mouth is positioned in front of the center microphone (No. 61). As already mentioned, the Minimum Energy Method is a near-field method, and therefore the center microphone was placed 3 cm in front of the mouth.

We would like to indicate that Takada and Bader have already done research on the radiation energy of singing voices in the same way [15]-[16]. However, the former was just a study on the radiation energy of singing at a fundamental frequency of 250 Hz. For the latter, we examined sound radiation of singing at other fundamental frequencies as well, i.e., like this study. But, only research findings from first analyses were displayed in [16], so that for this paper, further and more particularizing analyses were executed and their results will be shown in the following.

4 RESULTS

The data gained from the measurement was analyzed from various points of view and displayed in different forms, so that the research findings will be shown in two ways. But before starting the discussions, we would like to introduce some abbreviated terms, because of a lack of space in the lists :

- left and right clavicle = LC and RC
- nasal bone = N.Bone
- left and right corner of the mouth = LCoM and RCoM
- left and right cheek = LCh and RCh
- left and right lower eyelids = LLE and RLE

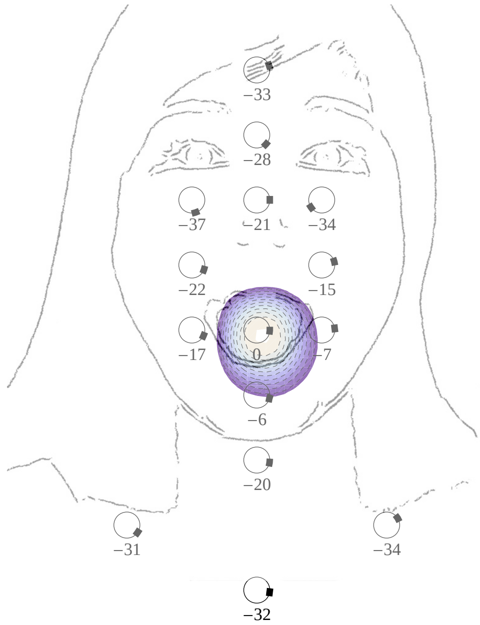


Figure 1: Radiation of a classical soprano (subject: SE1) at a fundamental frequency of 250 Hz for the vowel /a/.

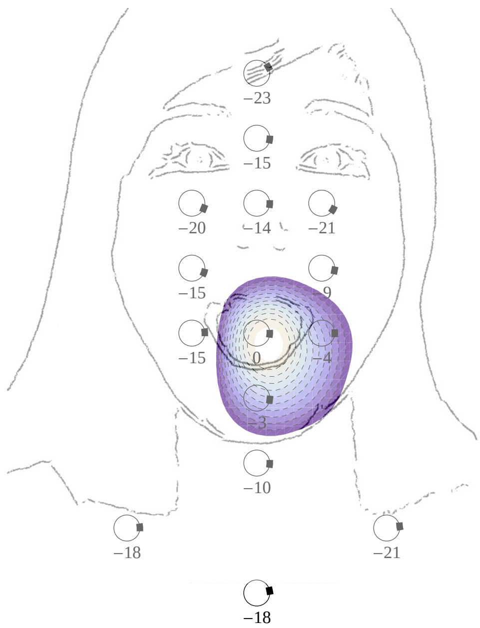


Figure 2: Radiation of a classical soprano (subject: SE1) at a fundamental frequency of 500 Hz for the vowel /a/.

Table 1: Process of the sound energy of a female musical theater singing (subject: SS) at a fundamental frequency of 250 Hz (in dB). The energy values shown in blue increase or decrease more than 10 dB compared to the energy value at the previous or subsequent frequency.

	Chin	Throat	LC	RC	Sternum	Nose	N.Bone	LCoM	RCoM	LCh	RCh	Forehead	LLE	RLE
/a/														
250 Hz	-18	-37	-52	-59	-43	-35	-35	-22	-29	-27	-34	-51	-31	-48
1000 Hz	-13	-29	-38	-33	-36	-27	-29	-20	-15	-23	-43	-39	-28	-40
2000 Hz	-12	-26	-25	-30	-31	-22	-26	-15	-21	-21	-29	-32	-25	-35
3000 Hz	-9	-20	-19	-23	-23	-18	-19	-8	-15	-13	-21	-21	-26	-26
4000 Hz	-12	-22	-25	-29	-26	-22	-23	-8	-12	-12	-12	-21	-26	-19
/e/														
250 Hz	-20	-38	-51	-51	-41	-33	-34	-27	-16	-28	-24	-47	-30	-47
1000 Hz	-15	-30	-37	-34	-34	-28	-27	-19	-7	-20	-21	-34	-26	-36
2000 Hz	-14	-24	-25	-29	-30	-22	-27	-16	-12	-17	-29	-32	-25	-32
3000 Hz	-14	-22	-20	-24	-23	-19	-21	-12	-14	-14	-26	-21	-25	-29
4000 Hz	-11	-21	-27	-29	-26	-20	-23	-8	-15	-11	-13	-21	-29	-21
/i/														
250 Hz	-32	-39	-52	-55	-44	-39	-37	-26	-17	-40	-29	-49	-34	-54
1000 Hz	-27	-33	-41	-47	-41	-33	-36	-25	-11	-32	-25	-36	-30	-34
2000 Hz	-24	-26	-32	-34	-34	-25	-29	-17	-11	-24	-33	-37	-27	-30
3000 Hz	-15	-22	-22	-24	-24	-20	-21	-13	-11	-16	-29	-23	-25	-29
4000 Hz	-14	-24	-26	-28	-31	-22	-24	-8	-13	-13	-16	-20	-25	-21
/o/														
250 Hz	-26	-52	-54	-57	-48	-42	-42	-24	-22	-42	-44	-56	-39	-52
1000 Hz	-23	-34	-43	-40	-42	-34	-36	-22	-15	-35	-34	-44	-33	-47
2000 Hz	-5	-15	-22	-21	-22	-15	-25	-6	-4	-11	-13	-27	-15	-18
3000 Hz	-12	-22	-21	-25	-25	-21	-20	-11	-8	-19	-26	-24	-25	-27
4000 Hz	-22	-27	-27	-30	-31	-24	-28	-15	-14	-18	-18	-24	-27	-22
/u/														
250 Hz	-23	-53	-53	-63	-47	-38	-41	-22	-24	-37	-43	-54	-38	-48
1000 Hz	-18	-34	-40	-41	-40	-33	-37	-20	-14	-30	-29	-44	-34	-42
2000 Hz	-8	-26	-29	-20	-25	-21	-26	-7	-6	-16	-16	-37	-15	-20
3000 Hz	-13	-20	-23	-25	-26	-20	-21	-14	-8	-18	-27	-23	-21	-25
4000 Hz	-16	-24	-24	-21	-27	-23	-26	-13	-16	-19	-18	-28	-23	-21

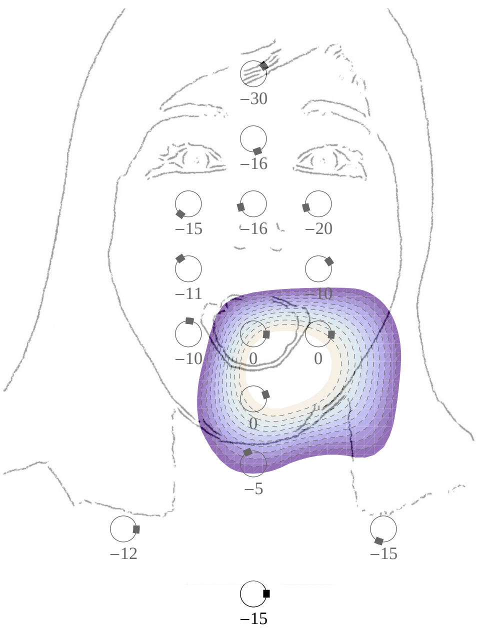


Figure 3: Radiation of the 7th partial singing by a classical soprano (subject: SE1) at a fundamental frequency of 500 Hz for the vowel /a/.

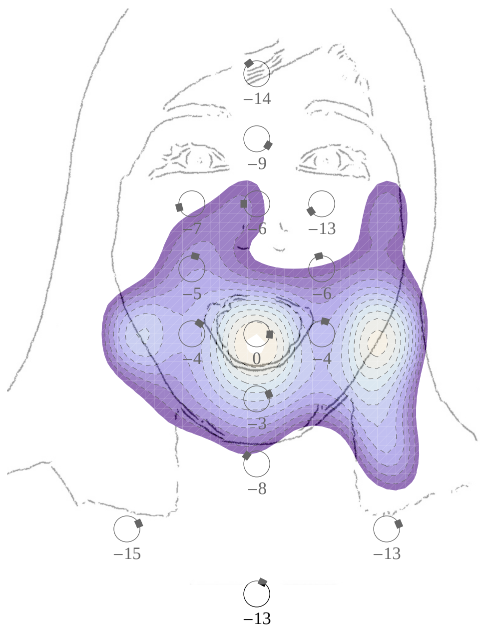


Figure 4: Radiation of the 10th partial singing by a female Pop singer (subject: SE2) at a fundamental frequency of 500 Hz for the vowel /e/.

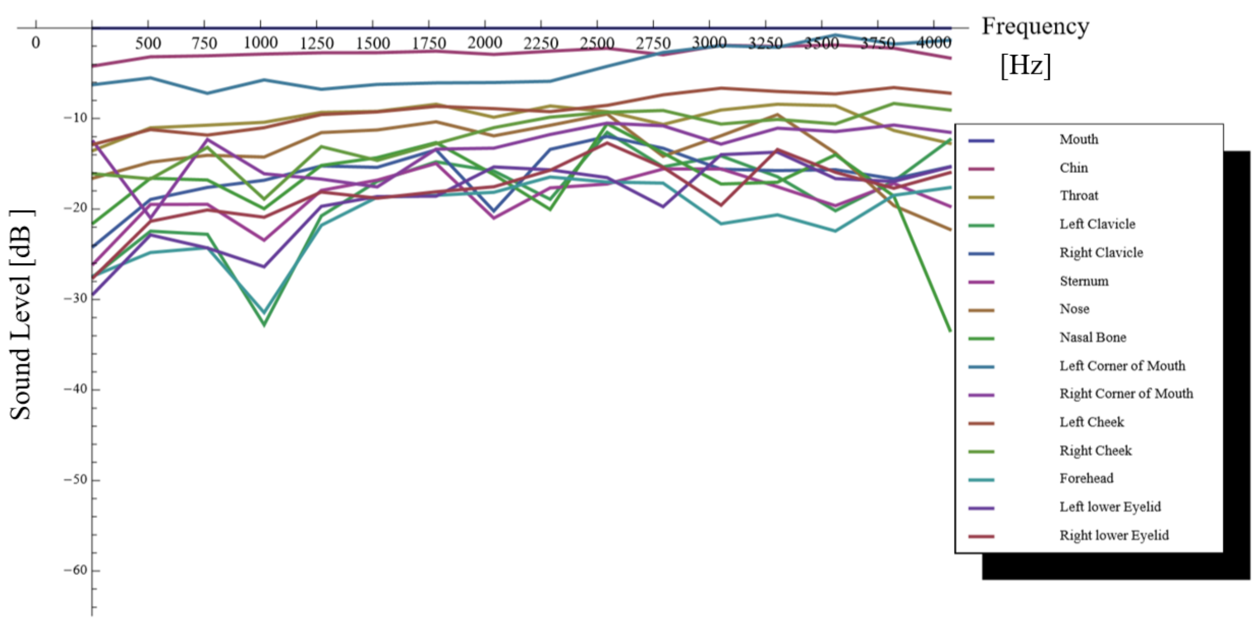


Figure 5: Sound levels of 14 measured parts of the body sung by a classical soprano singer (subject: SE1) relating to the radiation from the mouth (0 dB) at a fundamental frequency of 250 Hz for the vowel /a/ (color-coded sound levels of the parts from top to bottom shown at 250 Hz on the y-axis: chin / left corner of the mouth / right corner of the mouth / left cheek / throat / right cheek / nose / nasal bone / right clavicle / sternum / forehead / left clavicle / right lower eyelid / left lower eyelid).

4.1 Source of the Radiation Energy in Singing

At the fundamentals, it seems for all the subjects that the strongest radiation energy of their singing voice (color-coded area) came uniformly from the mouth, as shown in Figs. 1 and 2. It was already expected that the mouth is generally the strongest radiation source of the singing voice. In addition, the corners of the mouth often vibrated strongest of all the measured parts except the mouth due to their spatial closeness to the mouth, in fact independent of vowel. In general, it was revealed that the closer to the mouth the measured part of the body, the stronger the radiation energy.

Furthermore, when singing at higher pitches, the radiation energy increased in total in comparison to that revealed from singing at lower pitches, so that strong radiation energy came from a large area of the lower part of the face (see Fig. 2). Therefore, by means of visualizing the strongest radiating area up to -6 dB that was marked in color, it is clearly noticeable that the area of strong radiated energy shown at the fundamental became large with increasing pitch that was sung.

The finding that for singing at high pitch, the radiation of the singing voices became stronger at high frequencies, is presumably caused by the high lung volume and high-speed opening and closing of the vocal folds at the pitch, so that strong sound pressure will be produced by these factors. This is just what Sundberg asserted, although the body regions concerned are not confirmed by his recognition: very strong sound pressures in the vocal tract and in the trachea generate the phonatory vibrations in the skull, neck, and chest regions [17].

Because of the higher sound pressure caused by singing at high pitch, it was observed that the radiation energy from some parts of the body can be as strong as that from the mouth at higher partials, as shown in Figs. 3 and 4. Such radiation energy was observed by a shift or an enlargement of the strongest radiated point (radiator), thus from the mouth to other face / body regions, or emergence of multiple energy sources that were often found in the case of the fundamental frequency of 500 Hz, especially. This can be a temporary, but also as a continuing phenomenon.

The radiation patterns observed in our study revealed that the sound radiation changes depending on the frequency, and that the energy from the region of the body outside of the mouth increases strongly up to about 3 kHz, as shown in Figs. 5 and 7, and Table 1. The chin and the corners of the mouth were the parts showing the strongest radiation of energy of the 14 measured parts of the body (except for the mouth radiation), and the radiation from the chin often remained almost unchanged in the whole frequency range analyzed, so that the difference between its maximum and minimum was relatively small. However, it was found that the radiation energy of the parts of the body that are located far away from the mouth, increases dramatically and fluctuates strongly compared to the parts which are in the place near the mouth. This sometimes resulted in a difference of more than 20 dB between the maximum and minimum energies at these parts.

The strong radiation energy from the measured parts of the body at high frequencies is presumably related to the fact

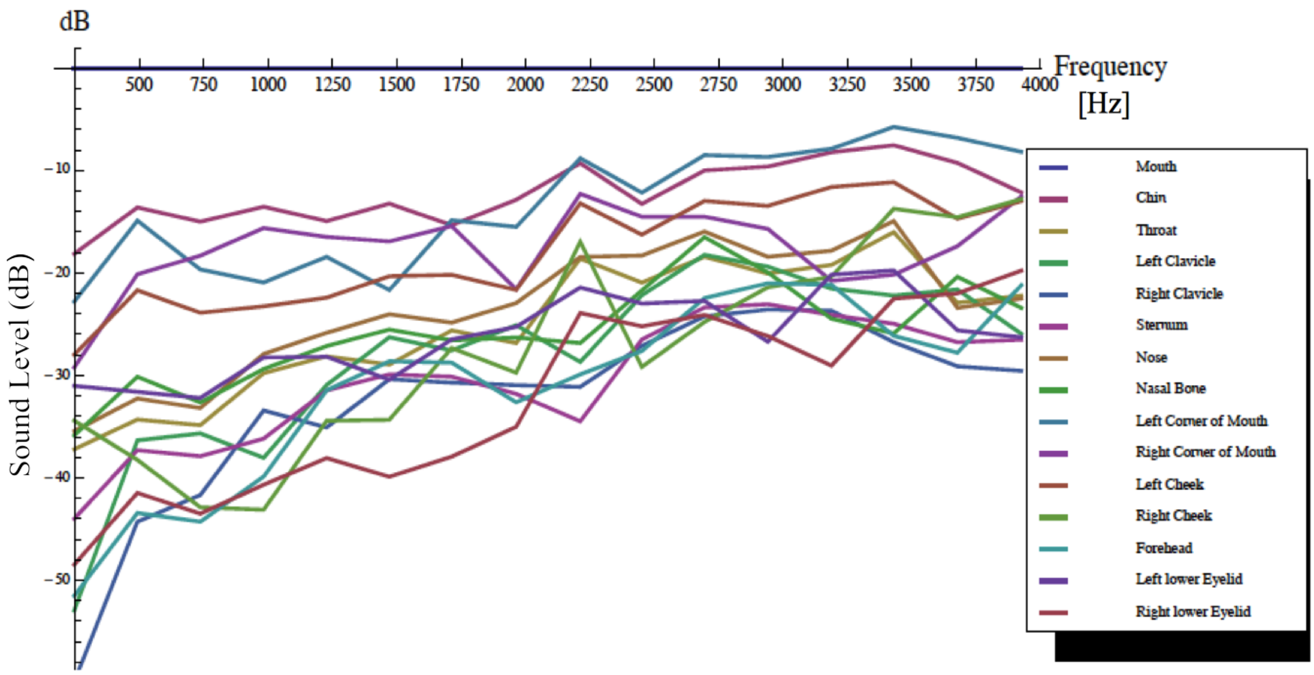


Figure 6: Sound levels of 14 measured parts of the body sung by a female musical theatre singer (subject: SS) relating to the radiation from the mouth (0 dB) at a fundamental frequency of 250 Hz for the vowel /a/ (color-coded sound levels of the parts from top to bottom shown at 250 Hz on the y-axis: chin / left corner of the mouth / left cheek / right corner of the mouth / left lower eyelid / right cheek / nasal bone / nose / throat / sternum / right lower eyelid / forehead / left clavicle / right clavicle).

that the partials at high frequencies radiate forwards, whereas these at low frequencies are emitted almost equally in all directions, as reported by Marshall and Meyer [18].

However, when it comes to the dimension of energy increase, there were large differences among the vowels and individuals. While the progress of the vowels /a/e/i/ was a smooth increase, that of the vowels /o/ and /u/ showed a strong zigzag-like course (see Figs. 5, 6 and 7, and Table 1. The energy from the mouth shown in both figures is displayed at zero on the x-axis (dark blue line).). This fluctuation was visible at all the parts of the body in many cases. Presumably, there were complicated air motions in the vocal tract due to the narrowing of the mouth opening and configuration changes of the vocal tract for these vowels. During the repeated recording and the analysis of both vowels, we heard and found a pronunciation with more air / noise in comparison to a pronunciation of the other recorded vowels, especially in case of nonclassical singers. But, only the results that are usable for this research were analyzed in this study, and therefore, such a zigzag-like course of both vowels is caused rather by complicated air motions in the vocal tract than by noises.

As mentioned above, the dimension of the energy increase was also different in each individual case. Some subjects showed a strong energy value from the beginning, i.e., from the fundamental frequency, so that their energy only slightly increased in comparison to the rest. For example, for classical singers, their energy levels gained from all the parts of the body were relatively close to each other for all the vowels, so that the difference of these energy levels was smaller among the measured parts in comparison to the nonclassical singers (see Figs. 5 and 7. The energy from the mouth is shown at zero on the x-axis (dark blue line).). Therefore, for the classical singers, there was no large increase of the energy from the region of the body outside of the mouth in the frequency range analyzed, in comparison to the energy of the nonclassical singers. This implies that the energy of classical singers radiates rather from their whole upper body while singing.

In regards to the difference in musical genre, the radiation energy of the nonclassical singing techniques for the vowels /o/ and /u/ strongly originates from the mouth area, i.e., from the corners of the mouth and the chin. This is clearly noticeable from the progress of the radiated energy, because there was a large difference in the energy level between these parts of the body and the rest of the parts measured, as shown in Fig. 7 (Top three lines are the results from the corners of the mouth and the chin). Presumably, this is due to different mouth/lip opening: classical singers usually hold the shape of their mouth/jaw opening as constant as possible for all vowels in order to keep certain vocal loudness and beauty of the voice, independently of the changing pronunciation of vowel sound. This difference in musical genre was observed distinctly at low pitch than at high pitch that is sung, likely due to the wider mouth opening at higher pitch.

In our study, we observed fluctuations of radiated energy of the singer's voice in the frequency range analyzed and ascertained that the radiated energy changes depending on the

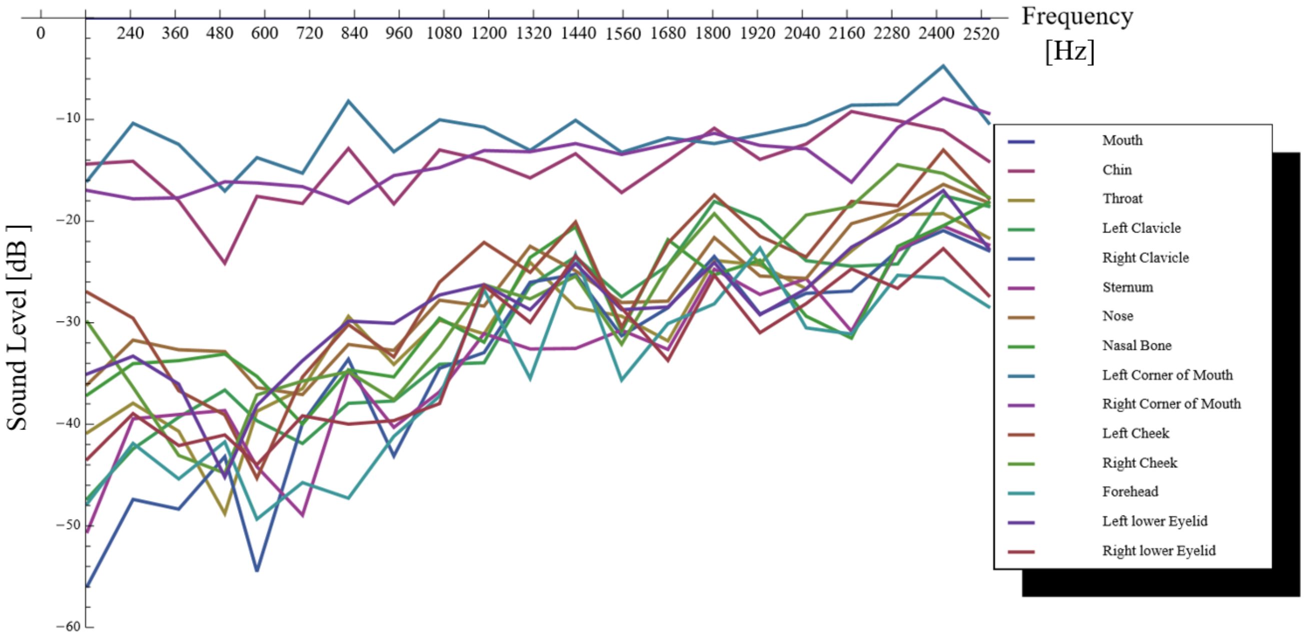


Figure 7: Sound levels of 14 measured parts of the body sung by a male musical theater singer (subject: TF) relating to the radiation from the mouth (0 dB) at a fundamental frequency of 120 Hz for the vowel /u/ (color-coded sound levels of the parts from top to bottom shown at 120 Hz on the y-axis: chin / left corner of the mouth / right corner of the mouth / left cheek / right cheek / left lower eyelid / nose / nasal bone / throat / right lower eyelid / left clavicle / forehead / sternum / right clavicle).

frequency. There it was shown where increasing and decreasing points are located in the frequency range, but because of the steady fluctuations, it was not so easy to realize the location of a minimum and maximum of the radiated energy. Therefore, the locations of these points of each involved subject – where these energies are located in the frequency range analyzed and how many participants showed their minimum and maximum energy there – are listed in Table 2. The numbers of the concerned participants are shown in red (minimum energy) and blue (maximum energy), respectively (there it is possible that there are several color-coded numbers in the tables, when multiple minimum or maximum result are therein. But in the case of spread results where all the subjects showed different results; i.e., there is neither a minimum nor a maximum number, none of the numbers are color-coded.).

This analysis revealed a rough relationship between sung pitches and the change of the frequency range where the minimum of the radiated energy was located, as the results from the vowel /a/ displayed in Table 2: in the cases of the fundamental frequency of 90 Hz (a single case study by a classical bass singer), 120 Hz, 180 Hz and 250 Hz, the minimum energy was located at the fundamental frequency, unlike the measurements at the fundamental frequency of 380 Hz (a single case study by a male musical theatre singer) and 500 Hz where this was not the case. In contrast, the maximum of the radiated energy was mostly located in the frequency range from 3 kHz and this was almost unchanged at each pitch, even though the frequency range analyzed varied according to the pitch that was sung. However, this localization differs slightly from vowel to vowel.

In other respects, the difference of the radiated energy amount among the singers revealed at the fundamental frequency becomes smaller at higher frequencies, even though a low energy value was observed at certain parts of the body at the fundamental frequency. For example, the sound levels of all measured parts of a classical soprano singer showed between ca. -4 and -30 dB (see Fig. 5) at the fundamental frequency of 250 Hz, whereas these of a female musical theatre singer were located between ca. -18 and -60 dB (see Fig. 6). But, at the 10th partial (2500 Hz), the difference of their sound levels is narrowed significantly, so between ca. -3 and -20 dB for the former and ca. -12 and -30 dB for the latter, respectively. Such a phenomenon is caused by a powerful energy increase of the subjects in the higher frequency range who actually revealed low energy values at their fundamental frequency (see also Fig. 7).

Furthermore, from the analysis of the minimum and maximum energy of the measured parts, it was found that there is a convergence of radiated energy values from all these parts in the frequency range between 2 kHz and 4 kHz (see Figs. 8 and 9, and Table 1). For example, in the case of the vowel /a/ sung by a female musical theater singer (SS) at the fundamental frequency of 250 Hz, the maximum and the minimum of the sound level indicated -18 (chin) and -59 dB (right clavicle) at 250 Hz, respectively (see Table 1). The difference of energy values from both parts was 41 dB. However, the difference between the minimum and the maximum of the sound level becomes smaller at higher frequencies (1000, 2000, 3000 and

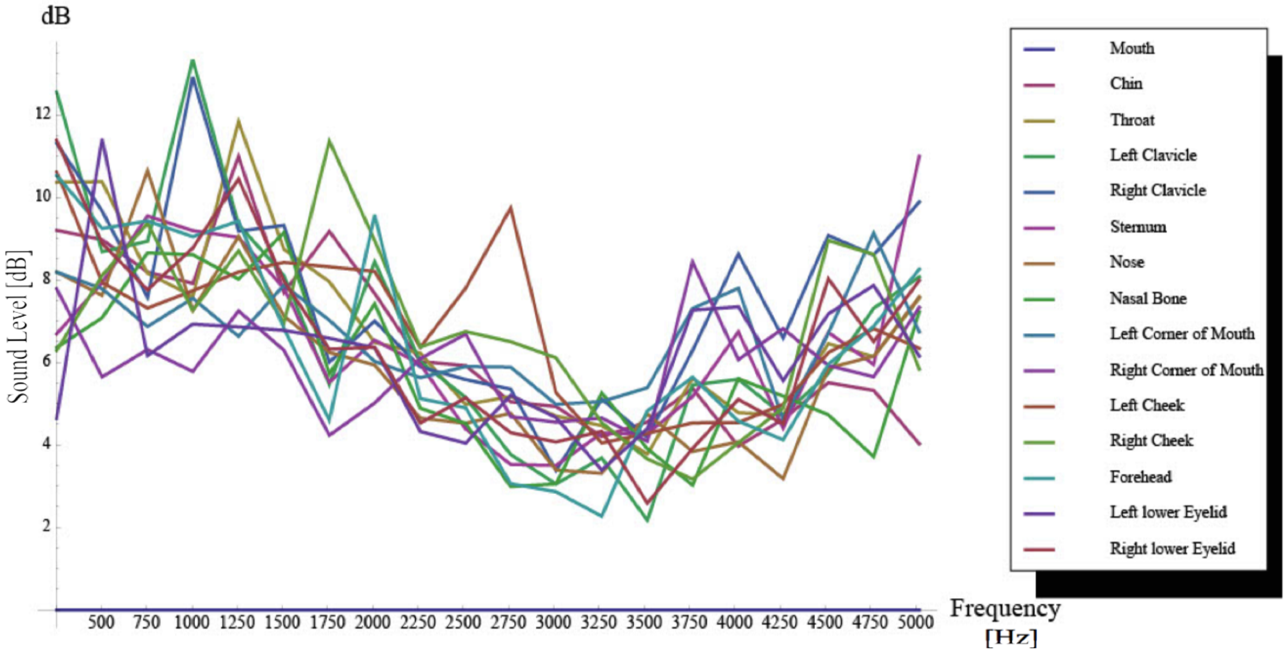


Figure 8: Difference of the radiation energy among the subjects, yielded from each part measured at a fundamental frequency of 250 Hz for the vowel /i/.

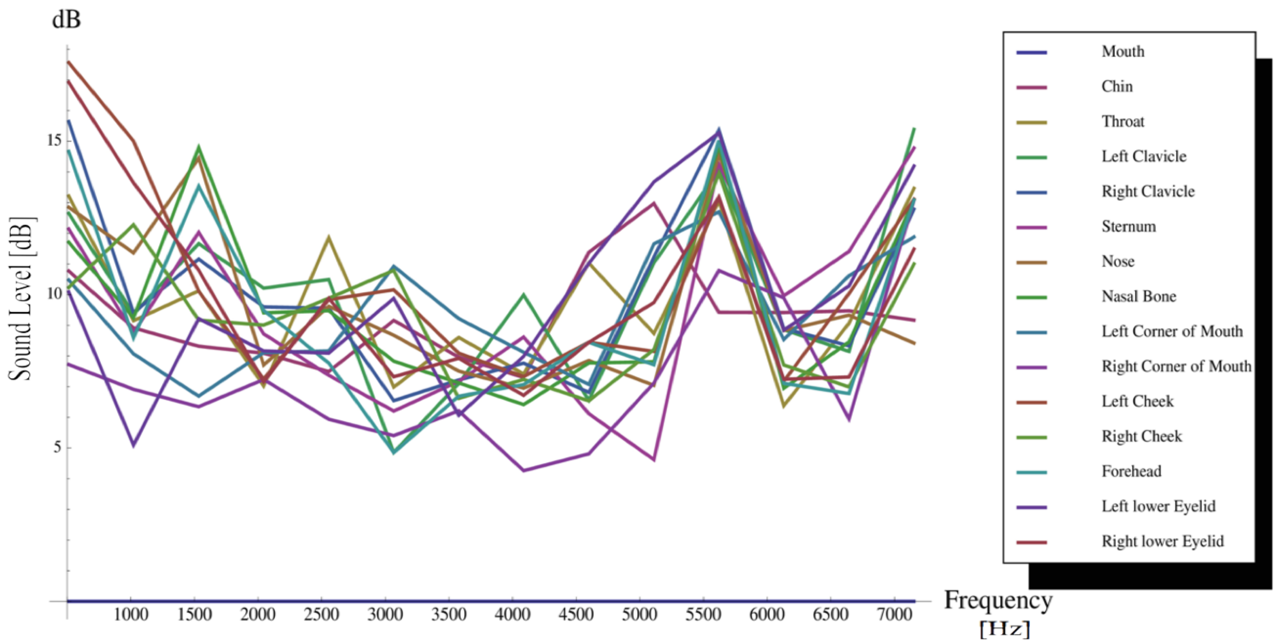


Figure 9: Difference of the radiation energy among the subjects, yielded from each part measured at a fundamental frequency of 500 Hz for the vowel /u/.

Table 2: Localization of the minimum (top) and maximum (bottom) of the radiated sound energy for the vowel /a/ singing at a fundamental frequency of 120, 180, 250, and 500 Hz shown by the number of participating subjects. The number of participating subjects (n) is displayed in the left column beside the sung fundamental frequency.

Localization of the minimum energy

	Chin	Throat	LC	RC	Sternum	Nose	N.Bone	LCoM	RCoM	LCh	RCh	Forehead	LLE	RLE
Fundamental 120 Hz, n = 2														
120 Hz	1	2	1	2	2	2	2	1	2	1	1	2	1	2
1000 Hz			1							1				
1500 Hz	1						1	1			1		1	
2000 Hz	1							1						
2500 Hz														
Fundamental 180 Hz, n = 4														
180 Hz	3	3	1	3	4	4	4	4	1	3	3	2	3	3
1000 Hz			2	1	1			2			2	1	1	
2000 Hz	1	1	1	1			1			1		1		1
2500 Hz									1				1	
3000 Hz									2					
Fundamental 250 Hz, n = 8														
250 Hz	7	7	4	7	8	7	6	6	3	7	5	7	7	7
1000 Hz			3					2	2	1	4	1		
2000 Hz	2	1	1					3	1	1		1		1
3000 Hz									1					
4000 Hz	1			1		2	2		1			1	1	
Fundamental 500 Hz, n = 6														
500 Hz	1	1		2		3			1	3	2	1	3	2
1000 Hz	1		5		3			3	2	2	2	3	2	1
2000 Hz	2	2	1	2	2	1	1	2	2	2	1	1	1	1
3000 Hz				1					1		1			2
4000 Hz	2	2		1	1	3	5		1					
5000 Hz	1	1					1	1				1		

Localization of the maximum energy

	Chin	Throat	LC	RC	Sternum	Nose	N.Bone	LCoM	RCoM	LCh	RCh	Forehead	LLE	RLE
Fundamental 120 Hz, n = 2														
120 Hz														
1000 Hz														
1500 Hz														
2000 Hz	1	1	1	1	1	1		1	1	1	1	1	1	1
2500 Hz	1	1	1	1	1	1	2	1	1	1	2	1	1	2
Fundamental 180 Hz, n = 4														
180 Hz		1							1	1		1		
1000 Hz									1				1	1
2000 Hz	2	2	2	1	1	1	1		1	1	2	1	2	1
2500 Hz		1	3	2		3	3		2		2	1	1	2
3000 Hz	3	3	2	1	3	1	2	4		2	1	1	1	1
Fundamental 250 Hz, n = 8														
250 Hz	1	1								1		1	1	1
1000 Hz								1	1					1
2000 Hz	2	1				2	1	1				1	1	
3000 Hz	4	6	3	6	7	6	6	4	2	4		5	4	1
4000 Hz	2	3	4	3	1	1	1	4	6	6	7	3	2	6
Fundamental 500 Hz, n = 6														
500 Hz			1				1		1	1		1	1	
1000 Hz	1			1					1					
2000 Hz	1	1	2			1	1					2		
3000 Hz	1		2	2	1		1	2		1	1			
4000 Hz	1	1	2					3	3	1	3	2		2
5000 Hz	6	4	1	3	5	5	4	3	1	4	2	2	5	5

4000 Hz), so that its difference was just 18 dB at 3000 Hz (the maximum: -8 dB at the left corner of the mouth, the minimum: -26 dB at both lower eyelids). The convergence of radiated energy values from all these parts in the frequency range can be seen clearly in Figs. 8 and 9. This finding shows that the difference of the radiated energy among the parts is smallest there due to its energy increase compared to the mouth radiation.

4.2 Total Radiation Energy from Measured Parts of the Body in Comparison to the Radiation Energy from the Mouth

When it comes to the sum of the radiated energy from all the measured parts of the body in relation to the mouth radiation, all the subjects showed that the total energy of their singing voice can exceed the energy from the mouth (see the upper diagram in Fig. 10. The result was compared to the mouth radiation, and thus, the line above zero point means that the total radiation energy is stronger than the radiation energy from the mouth.). But the energy level strongly depended on the frequency analyzed, the vowel and the pitch that was sung, just like the research findings shown in Subsection 4.1.

Comparing the radiated energy from the mouth with the total energy from all the parts excluding the energy from the corners of the mouth (i.e., energy from the following parts is included in this total energy: chin / throat / left and right clavicles / sternum / nose / nasal bone / left and right cheeks / forehead / left and right lower eyelids) made it clear that the energy from the corners of the mouth has a strong influence on the total energy of the singing voice (see the lower diagram in Fig. 10). Particularly, this influence seemed to be strong for the vowels /o/ and /u/. There was a large difference

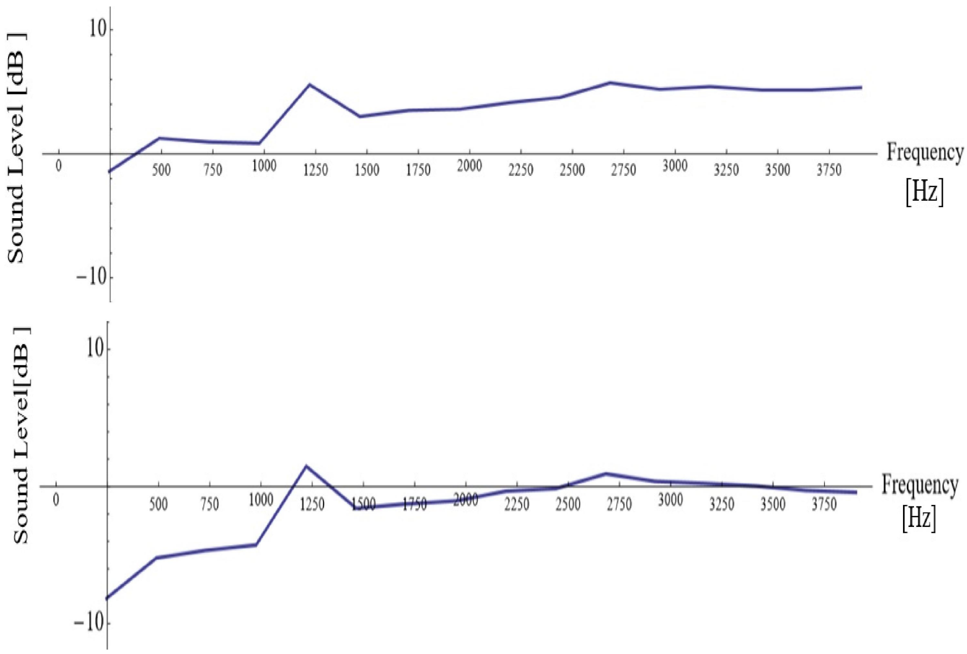


Figure 10: Total radiation energy from 14 measured parts of the body (top) and the total energy excluding the energy from the corners of the mouth (bottom) compared to the mouth radiation sung by a female Pop singer (subject: SE2) at a fundamental frequency of 250 Hz for the vowel /a/.

among subjects and independent of the musical genre. This recognizable difference was distinctive in the progress of the radiated energy analyzed at three partials: at the fundamental and at the partials where the maximum and minimum energy was measured (see Table 3).

But it was also shown that the total energy exclusive of the radiation energy from the corners of the mouth can still exceed the energy from the mouth in general. This emerged either as a temporary occurrence or as a continuing phenomenon at higher frequencies.

5 CONCLUSION

This study deals with the question of the radiated energy of the singing voice from the body in general. In order to find out whether most of the radiation energy comes from the mouth or not, an experiment was undertaken with seven singers from four various musical genres, and their singing voices were measured at 15 parts of the body by an acoustic camera with 121 microphones. The data gained from the measurement was analyzed from various points of view and displayed in different forms, so we hope that some questions have been answered thereby.

Our research findings showed that the mouth is usually the strongest sound source in singing, as expected. However, sufficiently loud sound pressures from other parts of the body, at least from the 14 parts measured, were also observed in this investigation. In most cases, an increase of the energy from these parts was shown up to about 3 kHz for all the vowels, so that the difference of the radiated energy among the parts of the body became smaller at high frequencies, even though the energy level showed a rather changeable process at all the pitches for the vowels /o/ and /u/. This increase depended on the pitch that was sung. A comparative analysis of all the findings gained from different sung pitches revealed that the strongest energy of the singing voice is usually located at around 3 kHz, independent of the singing technique.

Because of this increase, in some cases, a few of the parts of the body revealed even high energy values as much as the mouth radiation at higher partials. In addition, we found that the total energy from the 14 parts even surpasses the radiation energy values from the mouth. This depended strongly on the frequency, but this fact is a remarkable finding for a better understanding about the origin of radiation energy of the singing voice.

Furthermore, it was clarified that the energy from the corners of the mouth has strong influence on the total radiated sound energy level by comparing the energy from the mouth with the total energy from the other parts excluding the energy from the corners of the mouth. This influence was expected, due to their spatial closeness to the mouth.

In other aspects, it was found that the total radiation energy of the singing voice is strongly supported by the enhanced energy from the body regions that are located far from the mouth, because a dramatic increase in energy was found

Table 3: Total radiation energy measured at the fundamental frequency (FF), the minimum (Min) and maximum (Max) energy points from both analyses including (with CoM) and excluding the energy from the corners of the mouth (without CoM) as well as the differences at a fundamental frequency of 250 Hz (shown in dB, the results of JR for the vowel /o/ and VS (Pop) for the vowel /u/ are eliminated due to a lack of recording quality).

		With CoM			Without CoM			Difference		
Subject	**Vowel**	**FF**	**Min**	**Max**	**FF**	**Min**	**Max**	**FF**	**Min**	**Max**
CH (classical, female, alto)	**a**	-2	-8	3	-9	-19	-4	7	11	7
	e	-3	-4	3	-10	-12	-2	7	8	5
	i	-1	-9	1	-9	-17	-4	8	8	5
	o	-14	-18	1	-23	-30	-7	9	12	8
	u	-11	-27	1	-21	-37	-8	10	10	9
SE1 (classical, female, soprano)	**a**	3	3	10	-1	-1	5	4	4	5
	e	4	4	12	-1	-1	8	5	5	4
	i	4	4	11	0	0	7	4	4	4
	o	2	2	10	-3	-3	5	5	5	5
	u	-1	-11	4	-8	-27	-1	7	16	5
JR (classical, male, bass)	**a**	-6	-6	20	-13	-13	17	7	7	3
	e	-11	-15	10	-20	-23	5	9	8	5
	i	-7	-8	4	-15	-15	-1	8	7	5
	o	-	-	-	-	-	-	-	-	-
	u	-10	-12	4	-18	-19	-2	8	7	6
SS (musical theater, female, mezzo soprano)	**a**	-17	-17	2	-26	-26	-4	9	9	6
	e	-12	-14	2	-25	-27	-7	13	13	9
	i	-17	-18	0	-39	-39	-8	22	21	8
	o	-22	-22	17	-38	-38	14	16	16	3
	u	-22	-22	10	-36	-37	4	14	15	6
TF (musical theater, male, tenor)	**a**	0	0	13	-5	-5	8	5	5	5
	e	0	0	9	-7	-7	3	7	7	6
	i	-6	-6	10	-9	-10	3	3	4	7
	o	-6	-10	4	-18	-20	-4	12	10	8
	u	-2	-3	8	-18	-19	-2	16	16	10
SE2 (Pop, female, mezzo soprano)	**a**	-1	-1	5	-8	-8	1	7	7	4
	e	1	-1	9	-6	-6	6	7	5	3
	i	0	0	10	-6	-6	7	6	6	3
	o	0	0	13	-10	-10	11	10	10	2
	u	-7	-9	3	-17	-19	-1	10	10	4
VS (Pop, female, mezzo soprano)	**a**	-1	-1	8	-9	-9	3	8	8	5
	e	-4	-4	4	-13	-13	1	9	9	3
	i	-8	-9	4	-15	-15	0	7	6	4
	o	-5	-5	10	-19	-19	6	14	14	4
	u	-	-	-	-	-	-	-	-	-
VS (Soul, female, mezzo soprano)	**a**	-2	-2	8	-14	-14	1	12	12	7
	e	-6	-6	8	-19	-19	4	13	13	4
	i	-11	-11	1	-21	-22	-4	10	11	5
	o	-3	-3	8	-21	-21	1	18	18	7
	u	-11	-11	6	-38	-38	0	27	27	6

there, and because the energy values from the body regions of the chin, the corners of the mouth, and the cheeks did not change as much. However, for the vowels /o/ and /u/, the radiation energy of nonclassical singers is strongly supported by the corners of the mouth and the chin. For the nonclassical singers, the difference in the radiation energy level between the region of the mouth (i.e. the corners of the mouth and the chin) and the rest of measured parts of the body was much larger than for classical singers. This finding indicated that classical singers use rather their whole body as a musical instrument, and therefore the radiation energy of their singing voice is emitted steadily from there, in fact, independent of the vowel, for example.

In conclusion, the results of this study revealed that the radiation energy of the singing voice is emitted not only from the mouth, but also from other regions of the body. However, the radiation energy of the singing voice depends strongly on the vowels, frequency, pitch and person who sings. This means that the radiation of the singing voice depends rather on the singing technique of each singer than on the musical genre.

We hope that our research findings can give singers and vocal teachers of various musical genres helpful information for their vocal training and teaching.

REFERENCES

[1] I.R. Titze, "The human instrument, Scientific American", Nature Publishing Group, Vol. 298, No. 1, pp. 94–101 (2008).

[2] J. Kirikae, T. Sato, H. Oshima, and K. Nomoto, "Vibration of the body during phonation of the vowels", Revue de Laryngologie, Otologie, Rhinologie, Vol. 85, pp. 317–345 (1964).

[3] G. Fant, L. Nord, and P. Branderud, "A note on the vocal tract wall impedance, Speech Transmission Laboratory Quarterly progress and status report", Vol. 4, pp. 13–27 (1976).

[4] J. Sundberg, "Chest wall vibrations in singers", Journal of Speech, Language, and Hearing Research, Vol. 26, No. 3, pp. 329–340 (1983).

[5] Z. Pawlowski, R. Pawloczyk, and Z. Kraska, "Epiphysis vibrations of singers studied by holographic interferometry", Proc. SMAC-83, Royal Swedish Academy of Music, Vol. 28, pp. 37–60 (1983).

[6] A. Sakakura and H. Takahashi, "Body wall vibration in trained and untrained voices", Vocal Physiology: Voice Production, Mechanisms and Functions (Fujimura, O., ed.), Raven Press, New York, pp. 391–401 (1988).

[7] O. Takada, "Radiation Pattern of Classical Singing Style", Systematic Musicology: Empirical and The-

oretical Studies (Schneider, A., v. Ruschkowski, A., ed.), Hamburger Jahrbuch für Musikwissenschaft, Peter Lang, Frankfurt/M, Vol. 28, pp. 163–173 (2011).

[8] R. Bader, M. Münster, J. Richter, and H. Timm, "Microphone Array Measurements of Drums and Flutes", Musical Acoustics, Neurocognition and Psychology of Music (Bader, R., ed.), Hamburger Jahrbuch für Musikwissenschaft, Peter Lang, Frankfurt/M, Vol. 25, pp. 15–55 (2009).

[9] R. Bader, "Reconstruction of radiating sound fields using minimum energy method", The Journal of the Acoustical Society of America, Vol. 127, No. 1, pp. 300–308 (2010).

[10] R. Bader, "Microphone array", Springer Handbook of Acoustics, Springer, pp. 1179–1207 (2014).

[11] R. Bader, "Radiation characteristics of multiple and single sound hole vihuelas and a classical guitar", The Journal of the Acoustical Society of America, Vol. 131, No. 1, pp. 819–828 (2012).

[12] O. Abe, Sound Radiation of Singing Voices, PhD Thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky (2019).

[13] N. Plath, From Workshop to Concert Hall: Acoustic Observations on a Grand Piano under Construction, PhD Thesis, Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky (2019).

[14] T. Ziemer and N. Plath, "Microphone and Loudspeaker Array Signal Processing Steps towards a Radiation Keyboard for Authentic Samplers", https://www.mdpi.com/2076-3417/10/7/2333, Applied Sciences, Vol. 10, No. 7, 2333 (2020).

[15] O. Takada, "Sound Energy Measurement of Singing Voice on upper Parts of the Body: A Research in Classical, Pop, Soul and Musical Theatre Singing", ISMA 2014, Le Mans, pp. 445–451 (2014).

[16] O. Takada and R. Bader, "Radiation Patterns and Changing Radiation Energy of the Singing Voice", SIG Technical Reports, Vol. 2018-SLP-122, No. 14, pp. 1–4 (2018).

[17] J. Sundberg, "Phonatory vibrations in singers: A critical review", STL-QPSR, Vol. 32, No. 1, pp. 37–51 (1991).

[18] AH. Marshall and J. Meyer, "The directivity and auditory impressions of singers", Acta Acustica united with Acustica, S. Hirzel Verlag, Stuttgart, Vol. 58, No. 3, pp. 130–140 (1985).

(Received October 30, 2020)
(Accepted May 16, 2021)

Orie Abe Orie Abe is a musicologist with interests especially in the sociology of music and the acoustics of the singing voice. Additionally, she is a classical singer who has sung as a soloist at the Deutsches Nationaltheater und Staatskapelle Weimar in Germany, among other halls. She has studied vocal music at the Aichi Prefectural University of Fine Arts and Music (B.A.) in Japan and the Hamburger Konservatorium (postgraduate studies) in Germany. In addition, she has studied Systematic Musicology, Historic Musicology, and Japanology at the University of Hamburg (Magister Artium) in Germany where she has received her PhD in Systematic Musicology as well. Currently she is a special researcher at the Information Communication Media Laboratory in Japan.

Rolf Bader Rolf Bader studied Systematic Musicology, Physics, Ethnology, and Historic Musicology at the University of Hamburg where he obtained his PhD and Habilitation. He is Professor for Systematic Musicology at the Institute of Systematic Musicology, University of Hamburg since 2007. His major fields of research are Musical Acoustics and Musical Signal Processing, Musical Hardware and Software Development, Music Ethnology, Music Psychology, and Philosophy of Music. He published several books and papers about these topics. He was a visiting scholar at the Center for Computer Music and Research (CCRMA) at Stanford University 2005-6. He was also working as a professional musician, composer, and artist, running recording studios, working as a music journalist, leading exhibitions, and running a cinema. He conducted fieldwork as an Ethnomusicologist in Bali, Nepal, Thailand, Cambodia, Myanmar, Sri Lanka, China and India since 1999.

Regular Paper

Moving Object Detection Method for Moving Cameras Using Frames Subtraction Corrected by Optical Flow

Tsukasa Kudo[†]

[†]Faculty of Informatics, Shizuoka Institute of Science and Technology, Japan
kudo.tsukasa@sist.ac.jp

Abstract - With the progress of the Internet of Things (IoT), numerous videos have been recorded by various mobile cameras, such as in-vehicle and wearable cameras. Therefore, it has become necessary to organize target information automatically using image recognition. This often requires the extraction of the target area from the video frames by object detection as preprocessing. However, it often becomes difficult to detect a moving object efficiently in a video recorded by a moving camera depending on the video's environment, which can include complex backgrounds. In this study, I propose a method for extracting the target area by creating a subtraction of target images between adjacent frames. In this method, the background-position is aligned based on the displacement vector in the optical flow and then subtracted. Moreover, when the target areas are consecutively extracted from the video, I show its accuracy can be improved by comparing the extracted area size with its moving average.

Keywords: optical flow, moving camera, wearable camera, video, object detection, frame subtraction, IoT

1 INTRODUCTION

With the progress of the Internet of Things (IoT), various sensors are connected to the network, and a large amount of data is collected and analyzed. For example, regarding videos, a large number of cameras have been deployed and used for various purposes, such as monitoring traffic conditions and the insides of buildings. This has led to a rapid increase in the number of videos that need to be processed. Therefore, it has become necessary to extract target information automatically from such a vast amount of video data.

On the other hand, deep learning's effectiveness in image recognition has been demonstrated [4], [9], [23] and applied to various fields, such as handwritten character recognition and face recognition [5], [1]. Furthermore, it has been used for image recognition from a large number of videos, such as automatic target or object recognition and abnormality detection [16].

In this study, I have been attempting to automate inventory management in factories. Factory workers are equipped with wearable cameras and discriminate their current location and target objects using deep learning from images extracted from videos. Since this is for indoor use, backgrounds do not move but are instead diverse, including various kinds of walls and equipment. As a result, for the location, I showed that the training data could be collected efficiently, and the discrimination accuracy could be improved by continuous discrimination against the same target while comparing the results [11]. For the objects, they were held in hand and recorded by the cameras. However, when the objects were small, I found that the accuracy deteriorated due to the background's influence, especially in the case of a complex background.

The image is often preprocessed for deep learning to suppress the influence of the background, wherein a relatively small area including the target (hereinafter, target area) is extracted from the image. For example, in face recognition, the face area in an image is first extracted using Haar-like features then the face is recognized within this area [24]. Furthermore, various studies have been conducted on object detection in images, including video frames. However, I could not find an efficient method for the detection of a moving object from frames recorded by a moving camera, as mentioned above.

For this problem, I propose a method for extracting the target from backgrounds utilizing the optical flow in this study. The two frames are superimposed, and their background positions are matched based on the optical flow, then the subtraction between the frames is generated. As a result, the target area is extracted in this subtraction because the target's motion is different from the background. Along with this, I propose a method to identify the frame of the video, in which the target is observed, by utilizing the difference between the adjacent frames. I also show that the extraction accuracy of the target area can be improved by comparing its individual size with the moving average.

The remainder of this paper is organized as follows. Section 2 shows the motivation of this study and related works, and Sec. 3 proposes a target area extraction method based on the optical flow. Section 4 shows the implementation of this method in the experimental system, and Sec. 5 shows the experimental results. Section 7 concludes this paper.

2 MIOTIVATION AND RELATED WORKS

2.1 Motivation of This Study

I have been working on improving inventory management in a machinery factory, where various parts are stored in bulk containers. Since these inventory quantities cannot be counted visually from outside the container, their stock-taking is a heavy workload for the workers. For this problem, I showed the inventory quantity can be estimated with practical accuracy from the image of the bulk container by applying deep learning [10]. However, the next challenge was to find an efficient way to collect these images, since a factory usually houses more than one thousand bulk containers.

Consequently, I focused on the fact that inventory changes when workers replenish or ship the parts. I conceived to estimate stocks using images extracted from videos recorded by cameras that the workers wore. Firstly, it was necessary to detect a worker's approach to a bulk container and the handling of machine parts. I used images extracted automatically from videos to detect the former with a certain accuracy by using the deep learning model trained to distinguish the target room's entrance and equipment [11].

I noticed for the latter that the worker needed to hold the parts in his hand for the inventory work. In other words, as shown in Fig. 1 (a), if the object held in hand can be recognized, the target parts can be distinguished automatically. So, I collected various such images to train the model and evaluate the accuracy of distinguishing the target. As a result, I found that the accuracy deteriorated for small targets because of the background's influence.

Concretely, I evaluated the discrimination accuracy of three groups of objects (large, small, and thin) using a multi-class classification model of deep learning. Firstly, the photos taken in the same room were divided into training and test data. Then, the model was trained with the training data and evaluated with the test data. Its accuracy was 84.8%. Next, the model was evaluated with photos taken in another room. As a result, the accuracy was 100% for the large objects, which occupy more than half of each photo, but it deteriorated to 72.5% as a whole.

Preprocessing is usually performed for such problems to extract the target area from the image, and then the target is recognized by using this area image. Various methods using optical flow have been proposed to detect a moving object in a video image recorded by a moving camera [26]. The optical flow shows the displacement vector between a pixel in one frame and another frame's corresponding one. This applies to all frame pixels in the dense optical flow [7].

First, I performed an experiment in which the target was moved in front of the wearable camera. The target was then extracted based on the difference in the optical flow between the background and the target. As a result, the target could be extracted with high accuracy when the background was flat. However, I found a problem in the case shown in Fig. 1 (a), where the background was complex, and the target was flat. As shown in Fig. 1 (b), the target area was split, and it was difficult to extract it completely.

The process of this extraction is as follows. Figure 1 (c) shows the optical flow's displacement direction where brightness increases counterclockwise from zero (black) pointing to the right. Figure 1 (d) shows the normalized displacement distance, in which the higher the brightness, the larger the displacement distance. As shown in (c) and (d), the brightness distribution was not separated between the background and target. Furthermore, the target's flat part was not distinguished as the moving part, because both of its displacement direction and distance were almost zero (black).

Figure 1 (e) is a binarization of the brightness in (c), created by making 20% of the luminance range in the target area

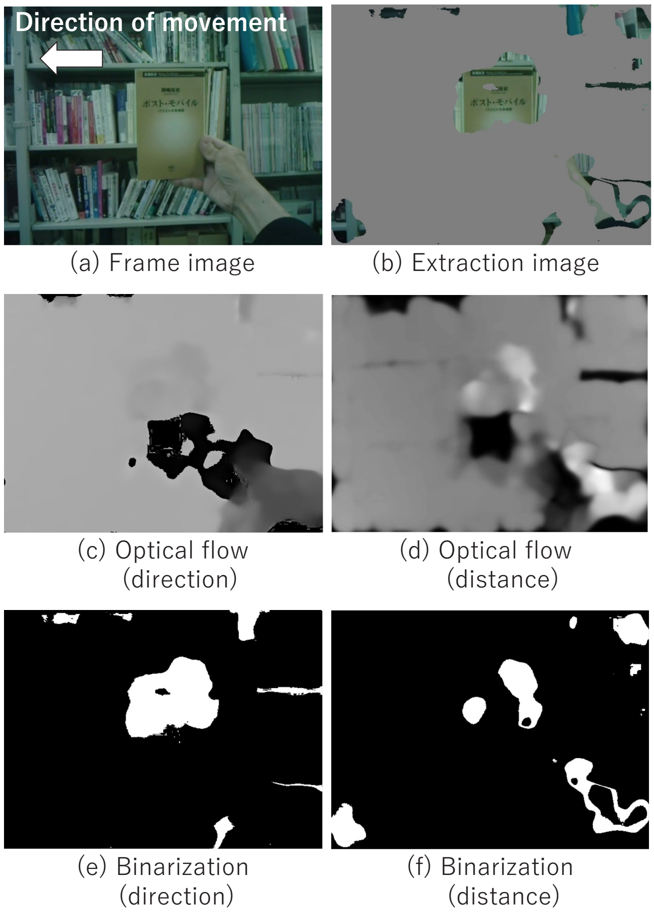


Figure 1: Problem of target extraction using optical flow

white. The central white area corresponds to the top part of the book in (a). Similarly, in Fig. (f), the white areas at the center and bottom right correspond to the book's top part and the hand in (a). These white areas were the target area estimated from the optical flow, and (b) was created by superimposing the original image on the combination of the white areas in (e) and (f).

This study's motivation is to develop an efficient method that can accurately extract the target area, including the moving target shot by a moving camera, even in the case of complex backgrounds. This study also targets stationary backgrounds and rigid objects that can be held and moved by the hand, as shown in Fig. 1 (a). Regarding the efficiency requirements mentioned above, the target areas must be consecutively extracted from the video, although there are some intervals.

2.2 Related Works

Currently, various kinds of mobile cameras have become ubiquitous, such as in-vehicle cameras and mobile phone cameras, in addition to the wearable cameras in this study. Therefore, many studies have been conducted to detect and track a moving object from videos recorded by mobile cameras.

The following three methods are widely used to extract a target from the frames of video: the background subtraction method, the frame subtraction method, and utilization of optical flow. In the background subtraction method, the target is extracted by the difference between the background image and the image in which the target appears in front of the background. In the frame subtraction method, the target is extracted using the subtraction between each frame. However, in both cases, it is assumed that the background image is fixed, and it is difficult to apply directly to a moving camera [15].

On the other hand, various studies using optical flow have been conducted for free-moving cameras [26]. The most direct ways use the displacement difference in the optical flow between the background and the moving target [17], [25]. However, as shown in Fig. 1, this creates a problem in the case of a complex background.

Some methods have been proposed for this problem. For example, combining different methods, performing analysis over many frames, and utilizing deep learning [2], [8], [15]. However, there were issues in terms of efficiency, such as processing complexity and model training. Moreover, methods to estimate the camera motion by utilizing optical flow have been proposed [6], [21]. However, their aim was motion recognition. Furthermore, some methods utilize the optical flow to reconstruct the background, and the target is detected by the background subtraction [22], [27]. However, these methods target seamless backgrounds or pan-tilt-zoom cameras. Therefore, it is challenging to apply these methods to the wearable camera shooting the complex background shown in Fig. 1.

Regarding object detection and classification of detected objects, studies utilizing deep learning have been progressing rapidly. Faster R-CNN performed both of them in a lump by collective end-to-end training of both models [20], and YOLO executed them with a single neural network to improve efficiency [19]. Concerning different scale objects, SSD made it possible to process them collectively [14], and RetinaNet improved efficiency by introducing the Feature Pyramid Network (FPN) and improving the error function [12], [13]. Then, M2Det has further improved accuracy and efficiency by introducing the new FPN and error function [28]. However, since these methods target each image, they are not suitable for object detection targeted by this study, which detects the objects moving in front of the background using multiple frames of a video. Furthermore, it is necessary to prepare a large amount of training data and train the model to apply these deep learning methods to the individual target shown in Fig. 1.

In summary, an efficient method has not yet been proposed to extract an area that includes a moving target in front of a complex background from a video recorded by a moving camera.

3 PROPOSED METHOD

This study aims to extract a target area from frames of a video shot by a wearable camera, as shown in Sec. 2.1, by utilizing the optical flow. First, it is necessary to extract suitable frames to calculate the optical flow between them, namely, frame pairs without considerable blur and observing the same target. The proposed method determines those frame pairs based on the difference between adjacent frames. Figure 2 shows an example of the transition of the difference between adjacent frames.

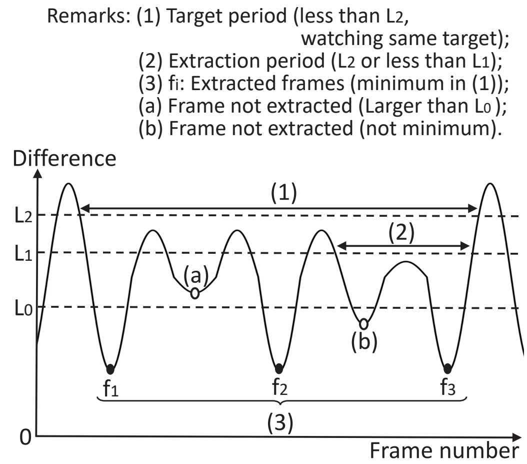


Figure 2: Transition of difference between adjacent video frames

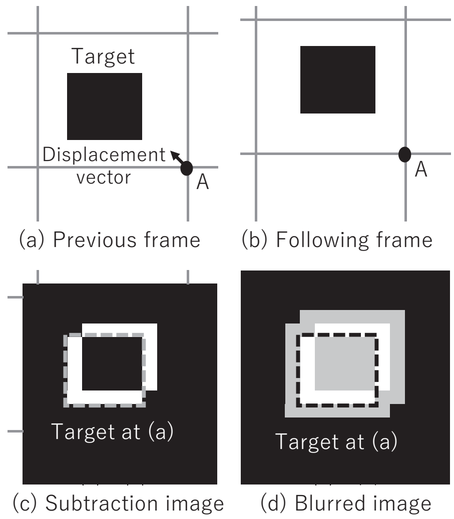


Figure 3: Target extraction method using optical flow and frame subtraction

Figure 2 (1) shows a target period surrounded by large difference points exceeding the threshold L_2. It is supposed the same target is continuously observed in this period. Incidentally, the viewpoint is supposed to have moved from one object to another at both ends. Only the frame with the least difference is then extracted for each extraction period shown in Fig. 2 (2) to obtain slightly different images for the same target. In the case of Fig. 2, they are f_1, f_2, and f_3. Here, the differences in the frames of (2) are L_1 or less, and the difference in each frame of (3) is L_0 or less to extract an image with little blur. The values of these thresholds are $L_0 < L_1 < L_2$. The frames (a) and (b) in Fig. 2 are not extracted because the difference of (a) is greater than L_0, and (b) is not minimum in (2).

Figure 3 shows the target area extraction process in the proposed method. (a) shows the video's previous frame f_1, and (b) shows the following frame f_2. Here, the black rectangle is the target. As shown in Sec. 2, the background is assumed to be stationary, so the difference in the background between (a) and (b) is only the parallel translation. Therefore, when the displacement vector of the optical flow between (a) and (b) is obtained for one point A in the background, (b) can be superimposed to (a) so that their background positions match. As a result, their subtraction in the background becomes zero (black), as shown in (c).

At this time, if the target was moved, there would be a gap in the target area between these two frames. This gap becomes the difference from the background, as shown in the white area in Fig. 3 (c). Note that there is no difference in the central area of the gap where the target overlaps in both frames, so it becomes black. Therefore, the target at (a), illustrated by the gray-dashed rectangle in (c), is included in this gap. Here, a part of the gap is outside this rectangle. However, since this study aims to narrow down the area where the target exists, this is acceptable.

As shown in the gray area in Fig. 3 (d), the image is blurred to enlarge the target area and connect the two white areas in (c); the entire interior of the area is also targeted. Further, some areas are set as no difference (black) to exclude the different areas between frames (a) and (b), which is caused by the displacement of (b). In the case of (c), they are at the left end and the top end (white). Finally, the target area can be extracted, which includes the target at (a) indicated by the black dashed line in (d) by extracting these white and gray areas as a continuous area.

4 IMPLEMENTATION

The functions described in Sec. 3 were implemented on a PC running Windows 10 to evaluate the proposed method. The programming language used was Python Ver.3.6; opencv-python Ver.4.1.0.25 was used for image processing. Below, the implementation of the proposed method is shown, along with two improvements.

4.1 Implementation of Proposed Method

First, the transition of the difference between the adjacent frames shown in Fig. 2 is calculated to extract target frames from a video. The frame image is converted to grayscale, and a histogram of the number of pixels with brightness j is created. This number is expressed by n_{ij} using the frame number i and brightness j. Then, as shown in Eq. (1), the absolute difference of D_i between previous and the following frames is calculated by weighting with luminance j and dividing by the number of pixels (N). Here, the division is to prevent fluctuations on D_i due to the number of pixels.

$$D_i = \sum_{j=0}^{255} |n_{i+1j} * j - n_{ij} * j| / N \tag{1}$$

D_i corresponds to "Difference" in Fig. 2. It is determined that the same target is observed while D_i is less than the threshold L_2. The frame with the smallest D_i is selected for each period where D_i is equal to or less than the threshold L_1. When it is equal to or less than the threshold L_0, this frame is extracted such as f_1 in Fig. 2.

The optical flow was calculated using the calcOpticalFlowFarneback method of opencv-python [18]. This is an implementation of the polynomial expansion algorithm [3]. In this method, I set the parameter as follows: the polynomial area was 5, the polygon width was 0.5, the window size was 60, the pyramid size was 0.5, its level was 3, and Gaussian kernel was used for prior blurring. Figure 4 shows an example of intermediate processing results of the proposed method for the same frame image shown in Fig. 1 (a). (1) and (2) show the displacement direction and distance of the optical flow. They are similar to Fig. 2 (b) and (c). Rectangles with a 10% margin from the edges are added for the following explanation. The background displacement is calculated using the median luminance of outside each rectangle. Note, the influence of the hand in the bottom right can be excluded by using the median.

If the target's displacement is too small, a clear subtraction image is not created, as shown in Fig. 3 (c). When the maximum displacement distance of the optical flow is less than the threshold, the subsequent frame is rejected, and the following one is tested. The threshold was set to 12 pixels based on the experimental results described later in Sec. 4.2. If the displacement was sufficient, the subsequent frame shifted based on the optical flow's background displacement, and its subtraction image with the previous frame is created. The distribution of luminance is then expanded by histogram equalization, as shown in Fig. 4 (3). If there is a difference between both frame regions due to the subsequent frame's shift, this part is filled with black, as shown at the left and bottom edges of (3).

The image is blurred by the median filter to combine image fragments as continuous regions, such as the book title, as shown in Fig. 3 (4). The kernel size of this median filter was set to 15×15 pixels. However, the corners of the targe are rounded by applying the median filter, as shown in (4). The measure for this is mentioned later in (7) convex hull and Sec. 4.3. (5) is a binarization image of (4) in which the area of the brightness above the threshold was extracted as white parts. This threshold was 159, which was the median value 127 plus the error 32. As a result, the part corresponding to the target was extracted as a continuous area, as shown in (5). Then, the white area was eroded and dilated to exclude noise and separate unnecessarily combined regions outside the target, as shown in (6). They were each performed five times with kernel size set to 3×3 pixels.

As shown in Fig. 3 (7), the target area is created by the following processing from (6). First, the largest continuous region of white parts is selected as the area, including the target. In the case of (6), the central region is applicable. Second, the entire part surrounded by the convex hull contour was extracted. Third, to recover the corners, which are rounded by the median filter shown in (4), the area is dilated. This was performed five times with the same kernel size as (6). Finally, the target region is extracted by superimposing the original image on (7), as shown in Fig. 3 (8). The part other than the target area is painted gray.

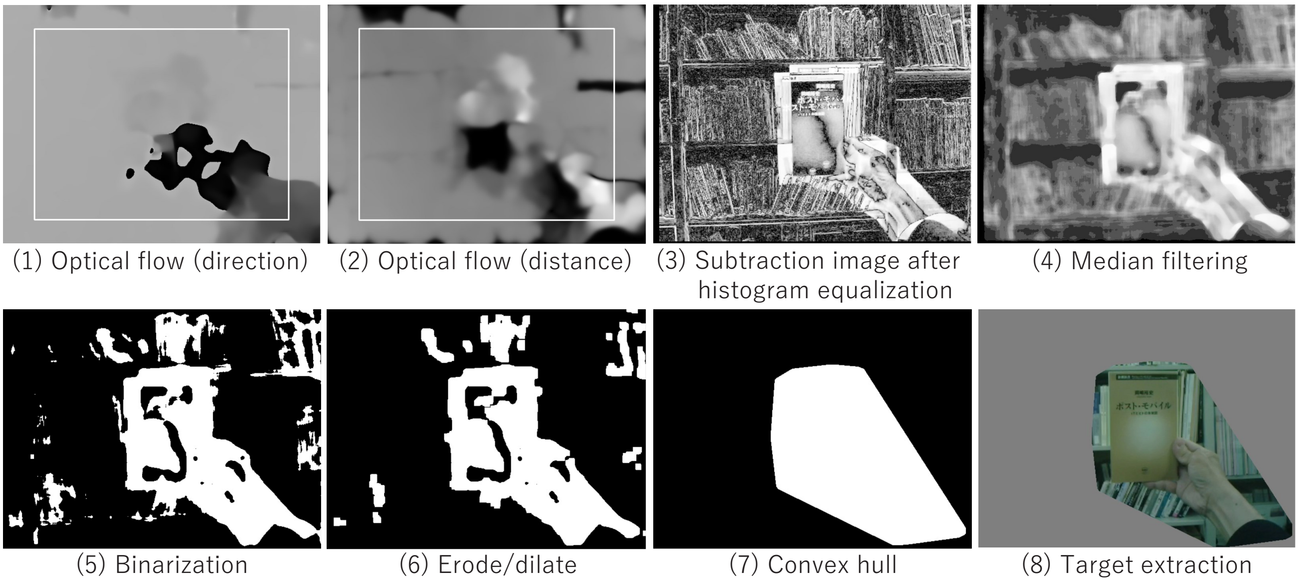


Figure 4: Intermediate processing result images in the proposed method

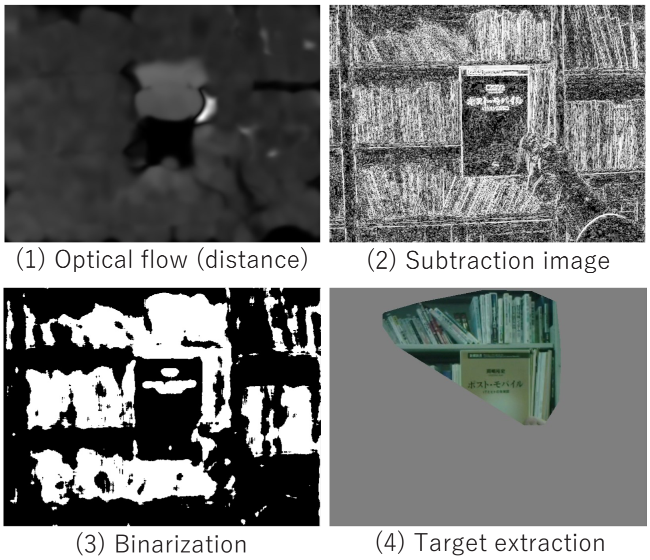


Figure 5: Analysis of partial missing of target area

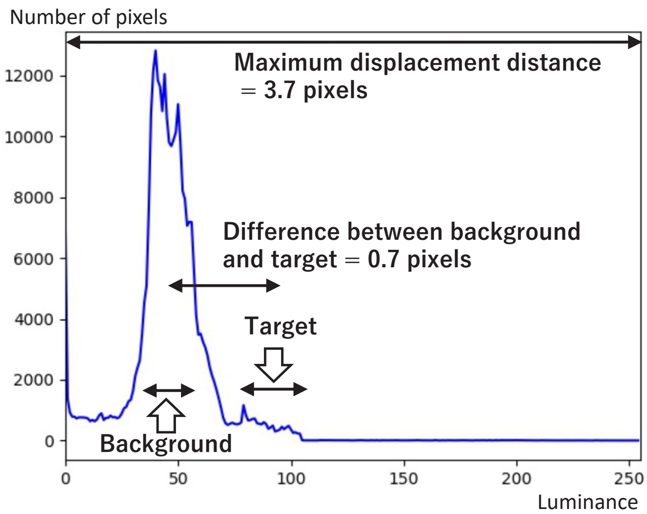


Figure 6: Histogram of distance in optical flow

4.2 Determination of Displacement Distance Threshold

As mentioned in Sec. 4.1, when the displacement of the target is too small, a clear subtraction image is not created, as shown in Fig. 3 (c). Figure 5 shows an example of such a case, and (1) shows the displacement distance of the optical flow. Figure 6 shows its histogram of luminance, and the maximum displacement distance is 3.7 pixels, which corresponds to a luminance of 255. The part with the maximum number of pixels corresponds to the background; the right part corresponds to the target. The difference in displacement distance between them is 0.7 pixels, that is, less than one pixel.

In this case, the luminance of the target subtraction image was small, and it was nearly the same as that of the background, as shown in Fig. 5 (2). Therefore, the background difference became too large in (3) binarization. As a result, the whole target area was not extracted, as shown in (4).

To find the displacement distance's threshold, I created simple images with a white background and a rectangle whose horizontal position changed stepwise. Then, the correlation between the displacement distances of the rectangle and the clarity of the subtraction image was evaluated. Besides, this image was created using the same procedure as in Fig. 5 (2). As a result, the difference between the target and background clearly appeared when the maximum displacement distance was 12 pixels or more. Based on this result, the threshold was set to 12 pixels in this implementation, as mentioned in Sec. 4.1

4.3 Restoration of Target's Corners

Figure 7 shows the cause and influence of the rounding of the target's corners by the median filter's blur shown in Fig.

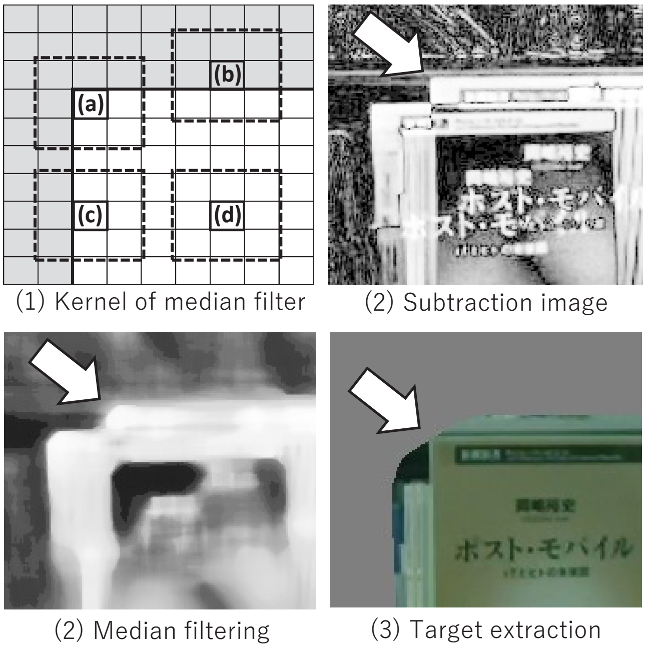


Figure 7: Rounded corners due to median filter

4 (4). Figure 7 (1) shows the target's boundary in the image. Each square corresponds to a pixel, and the white and gray areas correspond to the white and black areas of the image in Fig. 4 (3), respectively. Each broken line shows the median filter's kernel for the points (a) to (d), respectively, and the values of these points after blur are the median values of the broken line region. The value of the point (d), whose kernel is in the white area, and the boundary points (b) and (c) do not change even after blur. However, the kernel of the corner point (a) consists of five black pixels and four white pixels, so it becomes black by the median filter, and the corner is rounded.

Figure 7 (2) and (3) show the images before and after blur respectively, and the corner in (3) was rounded. As a result, in the final target extraction image, the corner was missing as shown in (3).

In this implementation, dilation of the convex hull shown in Fig. 4 (7) was performed eight times with a kernel of 3×3 pixels to prevent this missing of corners. This number corresponded to half of the median filter kernel size, which was 15×15 pixels. Note that dilation was also performed at the stage of Fig. 4 (6). However, if additional dilation was performed here, binding with fragments in the outside target area might occur. This process was performed in Fig. 4 (7).

5 EXPERIMENTS AND EVALUATIONS

The effectiveness of the proposed method was evaluated by performing the following two experiments. The first one aimed to evaluate the effectiveness in plural environments, and the extractions of the target area were performed for the combinations of three targets and four backgrounds. The second experiment aimed to evaluate the achievement of this study's purpose. In the environment shown in Fig. 1, automatic target extraction was performed continuously for a certain period of time.

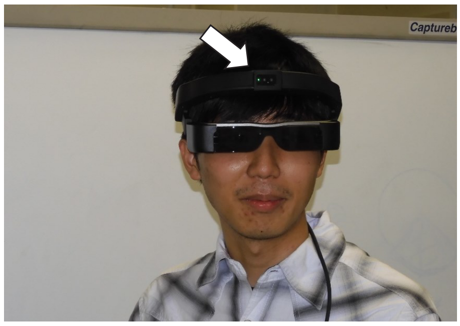


Figure 8: Wearable camera used in experiment

Table 1: Target object in experiment

No.	Type	Used target
B1	Clear contour	Book 1
B2	Clear figure	Book 2
B3	Flat	Book 3

A wearable camera recorded videos in a laboratory, for which headset EPSON MOVERIO Pro BT-2000B shown in Fig. 8 was used. It secured a video camera to the forehead, as shown by the arrow in Fig. 8. The video was displayed on the see-through glasses. It was set up with a frame size of 640×480 dpi at 30 frames per second. Images were extracted from videos shot by this camera using the experimental system mentioned in Sec. 4.

5.1 Evaluation in Plural Environments

For the experiment, I used three types of targets, namely books shown in Table 1, and four types of backgrounds shown in Table 2. The effectiveness of the proposed method was evaluated by combining these targets and backgrounds. As shown in Fig. 9, books consisted of the following: B1 had a clear outline in the bottom half, B2 had a clear form and an unclear outline, and B3 had a relatively flat image. Similarly, backgrounds consisted of the following: W1 was a flat wall; W2 was a relatively simple wall with equipment placed in front of it; W3 was a background with a clear boundary by the monitor; W4 was a complex background of the bookshelf. Note that the case of B3 and W4 shown in Fig. 9 is the one shown in Fig 1 (d).

First, a video of each target was shot with the background changing. The transition of difference between the adjacent frames was calculated using Eq. (1). Figure 10 shows the case

Table 2: Background in experiment

No.	Type	Used background
W1	Flat	Wall without equipment
W2	Sparse	Wall with equipment
W3	Bordered	Wall with large monitor
W4	Complex	Book shelf

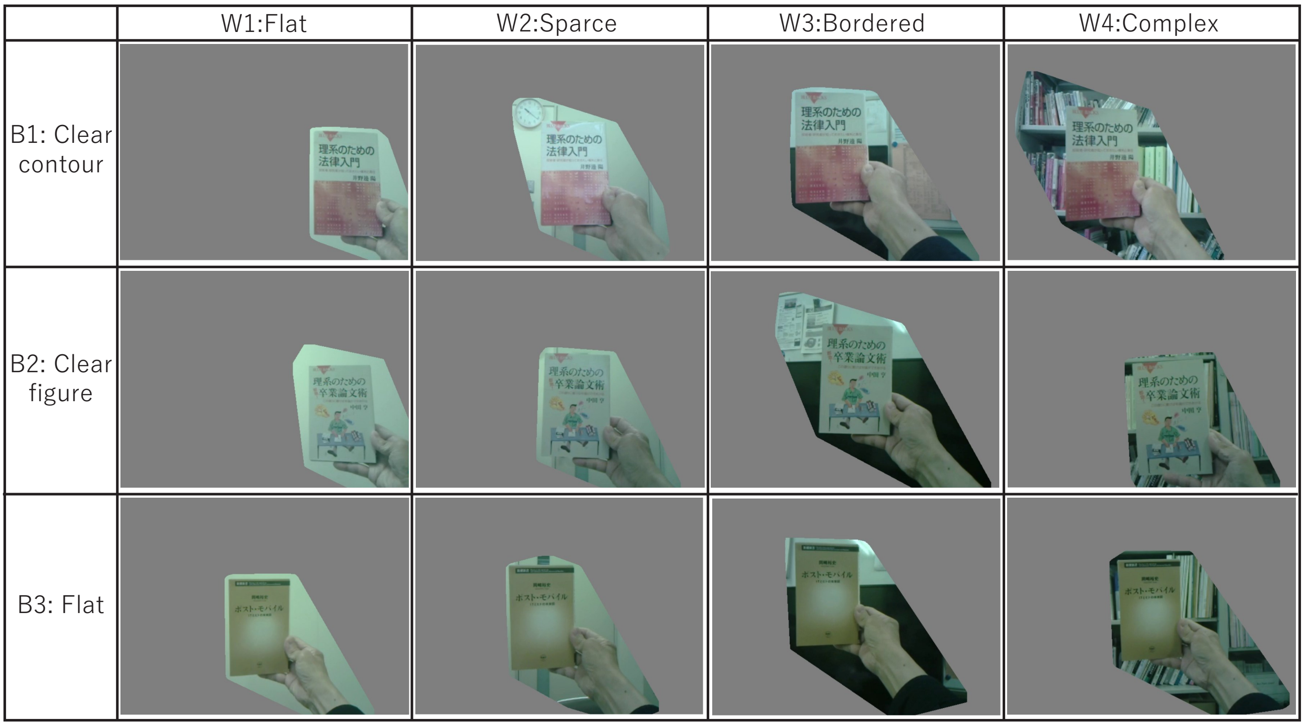


Figure 11: Experimental result in each combination between target object and background

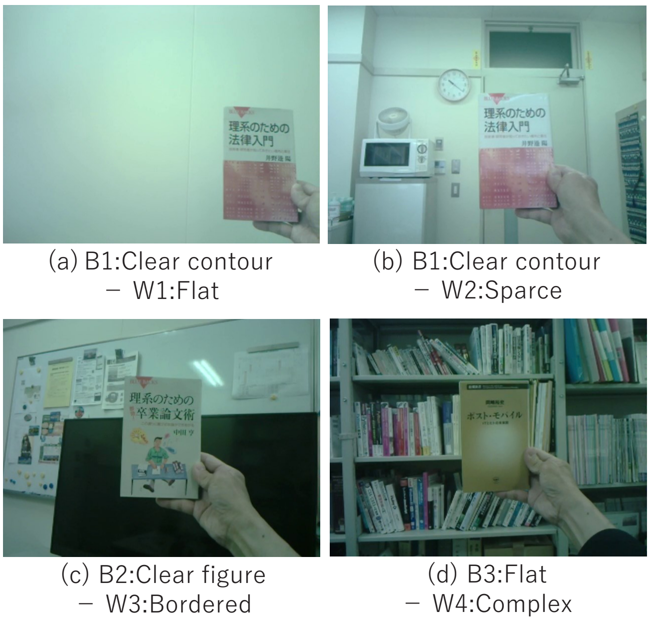


Figure 9: Combination examples of targets and backgrounds

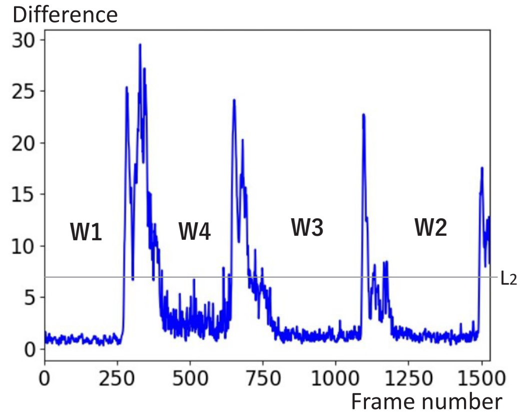


Figure 10: Transition of difference between adjacent frames as for target B1

of book B1. The relatively flat periods in Fig. 10 correspond to when the book was moved in front of each background, that is, the background was the same. The relatively large fluctuation corresponds to the camera's movement from one background to another. The movement could be detected based on the magnitude of the fluctuation, exceeding the threshold L_2. W1 to W4 in Fig. 10 correspond to each background, respectively.

I extracted a frame pair from these videos for each combination with a target displacement distance of 12 pixels or more based on the results described in Sec. 4.2. The target area was then extracted from each frame pair using the experimental system described in Sec. 4.1. Figure 11 shows the results where each row corresponds to the target, and each column corresponds to the background. The target area was extracted in every combination.

However, in combinations B1-W2, B1-W4, and B2-W3, areas other than the target were included. Furthermore, in the case of B2-W1, in which the background was flat, the top left part outside the target area was also extracted, although it was a relatively narrow range. Therefore, I investigated the intermediate results of the extraction process.

Regarding the former, Fig. 12 shows the case of B1-W4, where (1) shows the result of blur by the median filter, and

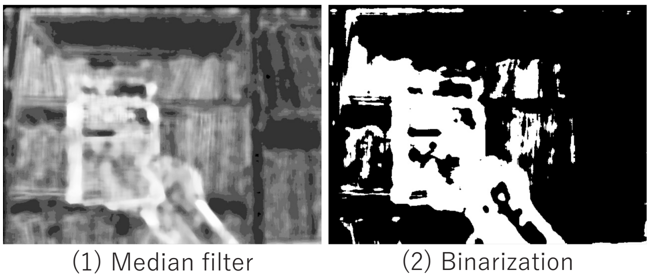


Figure 12: Intermediate processing result images in B1-W4

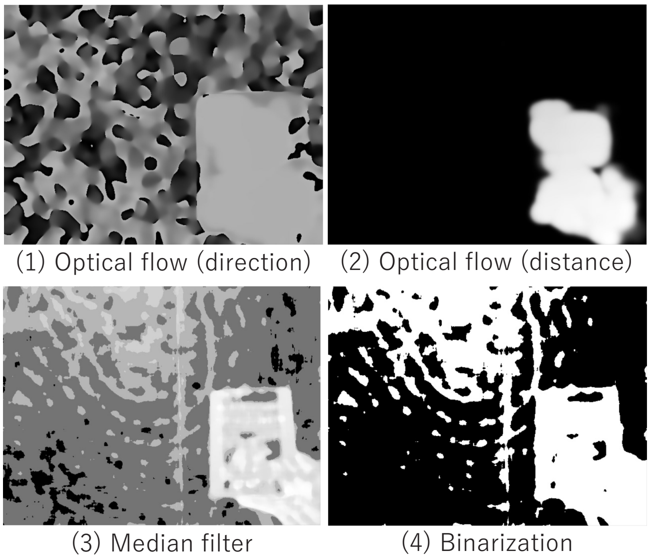


Figure 13: Intermediate processing result images in B2-W1

(2) shows the result of binarization. These correspond to (4) and (5) of Fig. 4, respectively. As shown in (1) of Fig. 12, the background subtraction was relatively bright on the left side and dark on the right. In other words, the distances of the displacement of the background were different depending on their position. Therefore, the upper left part of the background, where the distance was relatively large, was also extracted in binarization and connected to the target area, as shown in (2). As a result, this part was also extracted as the target area.

Regarding the latter, images of displacement direction and distance of optical flow are shown in Fig. 13 (1) and (2). They correspond to Fig. 4 (1) and (2), respectively, though the rectangles are not drawn. Additionally, Fig. 13 (3) and (4) is the same as Fig. 12 (1) and (2). This case used background B1, namely, the flat wall, so the background's displacement distance due to optical flow was calculated as zero, as shown in (2). However, there was a background difference between frames, as shown in (3). As a result, similar to the former case, the upper left part of the background was extracted in binarization and connected to the target. In other words, the displacement of the background could not be detected by the optical flow in the case of a flat wall.

However, the luminance of the background subtraction was smaller than that of the target, as shown in Fig. 13 (3). Therefore, it is expected that the background's influence can be suppressed by increasing the threshold of the binarization. This is evaluated in the next section.

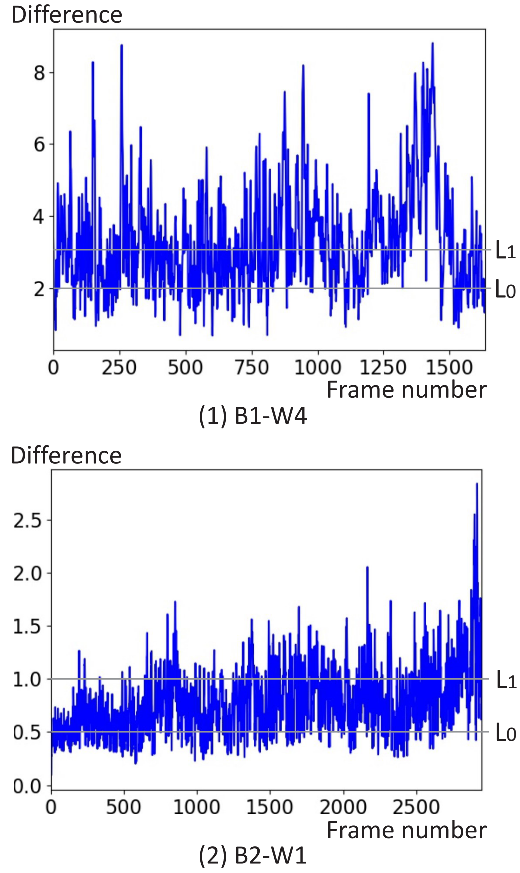


Figure 14: Transition of difference between adjacent frames of target period

5.2 Evaluations of Successive Extraction from Video

An experiment was performed to automatically extract the target area from the video and evaluate the accuracy for combination cases B1-W4 and B2-W1 shown in Fig. 11. A relatively wide area was extracted in these cases, as mentioned in Sec. 5.1.

First, the transition of the difference shown in Fig. 2. was graphed for each case. As shown in Fig. 14, the magnitude of the difference depends on the background. The magnitude was relatively larger in B1-W4, which had a complex background. Therefore, the threshold L_1 and maximum value L_0 were set to 3.0 and 2.0 in (1) B1-W4 and 1.0 and 0.5 in (2) B2-W1, respectively. As a result, 72 frames were extracted from about 1,630 frames of the video in (1), and 69 frames from about 2,950 frames in (2).

The target areas were then extracted using the adjacent pairs of extracted frames. Here, only the pairs with a maximum displacement distance of 12 pixels or more were used, similar to Sec. 5.1. Accordingly, each previous frame's subsequent frames were sequentially tested, and the first frame with 12 pixels or more pixels was selected as its pair. Two thresholds for the binarization of 159 and 191 were used in (2), which

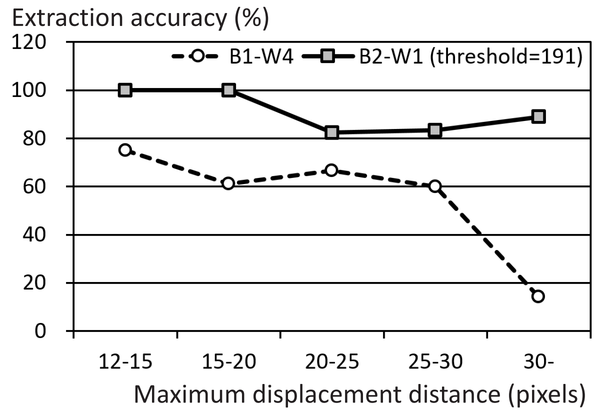


Figure 15: Correlation between accuracy and maximum displacement distance

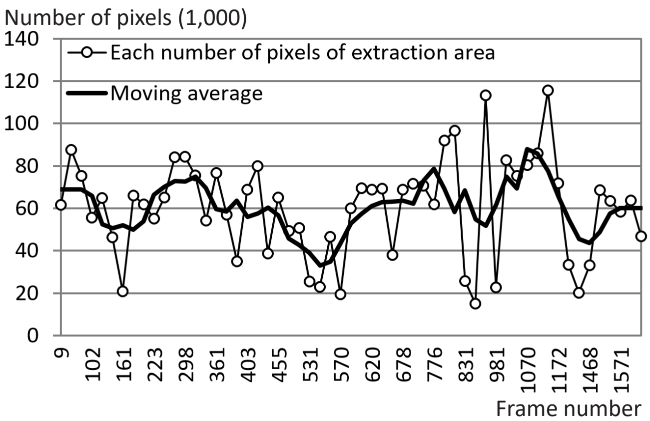


Figure 16: Transition of each number of pixels of extraction area and its moving average (window size=5)

had a flat background, as mentioned in Sec. 5.1. Whereas, only one of 159 was used in (1), similar to Sec. 5.1. Here, the error 32 is doubled for 191, namely, $191 = 127 + 32 \times 2$. The number and accuracy of target area extraction were 57 and 59.6% in B1-W4, and 59 and 88.1% in B2-W1 with the threshold 191; 59 and 100.0% in B2-W1 with the threshold 159.

The correlation between the extraction accuracy and displacement distance was evaluated to clarify the appropriate displacement distance range. Figure 15 shows that for B1-W4 with complex backgrounds, the accuracies were 60% or more in the cases with 30 pixels or less. However, the accuracies deteriorated to about 14% in cases with more than 30 pixels. On the other hand, in the case of B2-W1 with flat backgrounds and threshold 191, no significant deterioration was observed within the experimental range. In addition, the case of B2-W1 with threshold 159 was omitted, because all the accuracies were 100%.

In this method, since the target area is extracted from the video consecutively, it can be assumed that the target areas have almost the same size in the nearby frames. In other words, when the target area is missing or too large, it is expected that there is a certain gap in the ratio between each target area's number and their moving average (hereinafter,

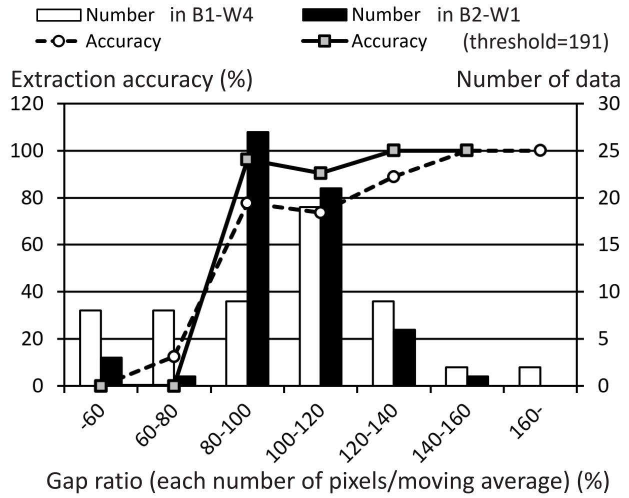


Figure 17: Correlation between gap ratio and extraction accuracy

gap ratio). Figure 16 shows the transition of each number of pixels of the extraction area in B1-W4 and its moving average with a window size of 5. The former fluctuates considerably, while the latter fluctuates gently. Regarding fluctuation factors of the moving average, there was the distance fluctuation between the camera and target, and the size fluctuation of the hand area, depending on the position of the target.

The correlation between the gap ratio and extraction accuracy was then evaluated. Figure 17 shows the results. The horizontal axis is the gap ratio, and when it is 100%, the size of the extraction area is equal to the moving average; The left vertical axis shows the extraction accuracy indicated by the line graph, and the right vertical axis shows the number of data indicated by the bar graph. B2-W1 is the case with a threshold of 191. The case of 159 is omitted because the accuracy was always 100%. As shown in Fig. 17, the extraction accuracies deteriorated significantly when the gap ratio was less than 80% in both cases. In the case of B1-W4, eight data with a maximum displacement distance of 30 pixels or more, shown in Fig. 15, were included; only one was not included.

Figure 18 shows examples of the extracted target areas in B1-W4. Both (a) and (b) show the cases where the gap ratios were around 1.0. While (a) is extracted without any parts missing, a part is missing in (b). Although (a) and (b) are close in the gap ratio, there is a large difference in the extracted area. This is because large or small extracted areas of nearby frames affected the moving average. Furthermore, (c) shows the case where the gap ratio is large, and a wide area was extracted. Conversely, (d) shows the case of a small gap ratio where a part is missing.

In B2-W1, there was a difference in the extraction accuracy, depending on the threshold. Therefore, the correlation between the size of the extraction area and the threshold was also evaluated. Figure 19 shows the result in a scatter plot. The vertical axis shows the number of pixels in the extraction area with threshold 159. The horizontal axis shows the case with threshold 191. Here, the data is excluded where a part was missing in the case of threshold 191. The diagonal line corresponds to the case where the number of pixels is equal in

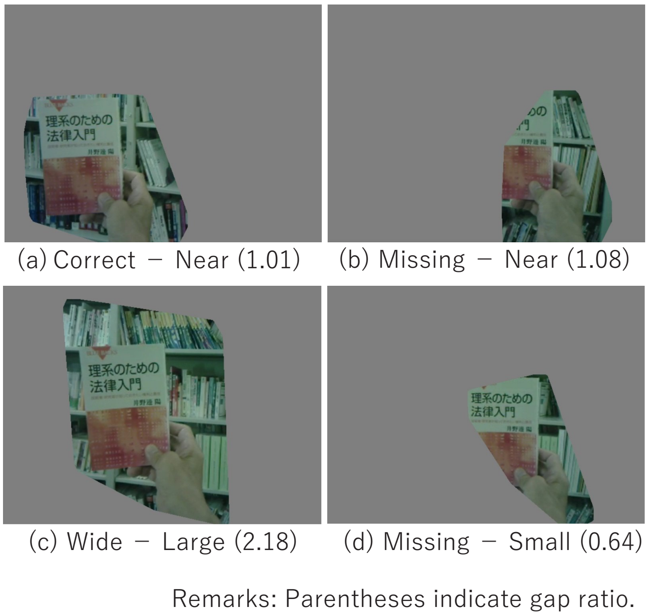


Figure 18: Examples of extraction results in B1-W4

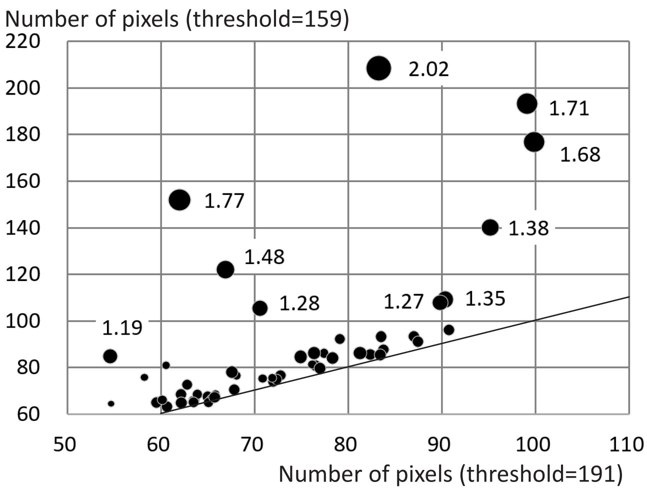


Figure 19: Correlation of the number of pixels of the extraction area between threshold 159 and 191 in B2-W1

both. The extraction area size of threshold 159 was equal to or more than that of threshold 191 in all cases. Furthermore, point sizes in the graph show the gap ratio of each extraction area of the threshold 159. The gap ratio increased when the deviation from the diagonal line increased. In other words, the size was very large compared to that of threshold 191.

Figure 20 shows an example of the extracted target area in B2-W1. The top row shows the case of threshold 191, and the bottom row shows that of threshold 159. The left column shows the case where extraction was accurate without missing parts or too-wide, and the two ((a), (c)) are the results for the same frame. As shown in (a) and (c), the smaller the threshold was, the larger the target area was. Also, similar to the complex background in B1-W4, missing occurred when the gap ratio was too small, as shown in (b), and a too-wide area was extracted when the gap ratio was too large, as shown in (d).

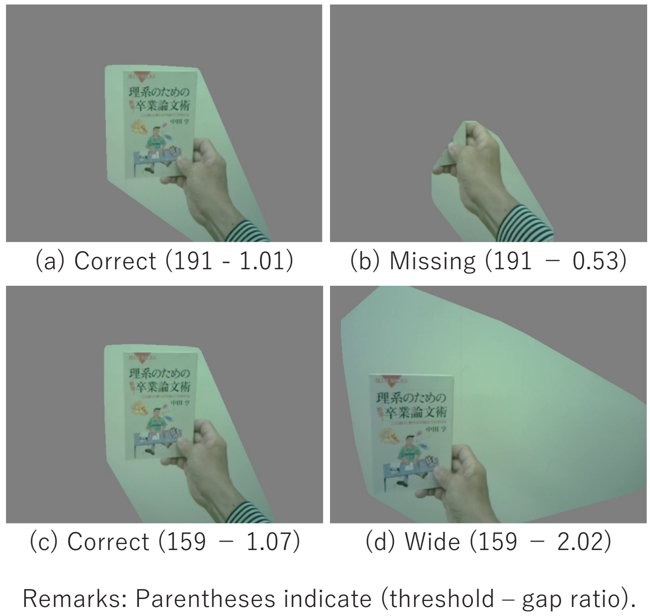


Figure 20: Examples of extraction results in B2-W1

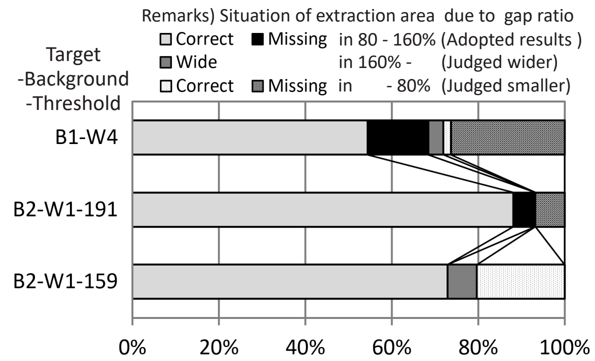


Figure 21: Percentage of extraction area situation due to gap ratio

The above results show that the extraction accuracy can be improved by adopting a gap ratio between 80% and 160%. Figure 21 shows the following percentages of accuracy: accurate and missing extractions when the gap ratio was between 80% and 160%; too-wide extraction defined the case where the ratio exceeded 160%; accurate and missing extractions when the ratio was less than 80%. Here, accurate extraction indicates the case without missing or too-wide extraction. In B1-W4 and B2-W1 with threshold 191, the missing ratio increased when the gap ratio was less than 80%. On the other hand, in B2-W1 with threshold 159, there were only the accurate extractions, even when the gap ratio was less than 80%. However, there were too-wide extractions.

Figure 22 shows a decrease in the error ratio when the adopted data were narrowed down based on the gap ratio, namely, between 80% and 160%. The error was defined as missing or too-wide extraction. In B1-W4 and B2-W1 with threshold 191, both error rates improved to less than half. In B2-W1 with threshold 159, the error rate also decreased to 0% by removing too-wide extraction. However, the data removal rate became about 27% because normal data were also

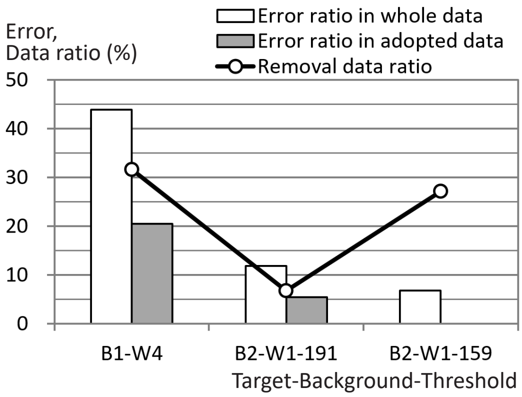


Figure 22: Improvement of error rate by narrowing data based on gap ratio

removed.

5.3 Evaluation Results

I proposed and evaluated a method for extracting a target area from a video shot by a wearable camera. The evaluation results are summarized below.

First, it was shown that the target area could be extracted with various backgrounds and targets, as shown in Fig. 11. In Fig. 11, an area surrounding the target was also extracted. As long as it is a narrow area, this is acceptable since this method aims to extract the area that includes the target from an image as a preprocessing of image recognition using deep learning. Besides, when the binarization's threshold increased, the ratio of missing parts from the extracted target area also increased, as shown in Fig. 21. When the threshold decreased, the relatively wide area was extracted.

Second, the target frames could be extracted from the video automatically and consecutively using the difference between adjacent frames. In this method, only the frames with small differences in which the object was observed were extracted. In other words, they had the maximum value (L_0) or less, as shown in Fig. 2. Therefore, a frame was not extracted when the difference between adjacent frames was greatly changed, such as during movement, as shown in Fig. 10. On the other hand, as shown in Fig. 14, the magnitude of the fluctuation of the difference between adjacent frames depended on the background. Therefore, the threshold (L_1) and maximum value (L_0) had to be set for each background.

Third, frames that might be missing or too-wide could be removed by comparing the size of the extracted target areas with their moving averages. Consequently, the target area could be extracted with an error rate of about 20% for a complex background, and about 5% or less for a flat background, as shown in Fig. 22. On the other hand, about 30% of the frames were removed in some cases of this improvement. However, this method's advantage is that a large number of frames could be easily collected since a video camera continuously recorded the target. For example, in B1-W4, 31 accurate target areas were automatically extracted from 1,630 frames, namely, about 54 seconds of video. In other words, one correct extraction image could be acquired every 1.7 seconds. Therefore, it is acceptable to reject a certain percentage of data in this method.

6 DISCUSSION

Firstly, I mention below that the comparison between the related works shown in Sec. 2.2 and the proposed method. Whereas the background subtraction and frame subtraction method targeted stationary cameras, the proposed method was able to extract the target even with the wearable camera, that is, the moving camera, shown in Fig. 8. Similarly, with the conventional method using optical flow, extracting the target was difficult in a complicated background as shown in Fig. 1. But, with the proposed method, it was possible as shown in Fig. 11. Therefore, as methods utilizing a few video frames of moving camera, I consider that the moving object detection accuracy could be improved by the proposed method.

In addition, comparing with the method combining different methods or analyzing many frames, the proposed method uses only two adjacent frames. So, for example, it is possible to repeatedly extract the target from the continuously shot video and improve the accuracy by comparing the moving average as shown in Fig. 16. Similarly, whereas applying deep learning to a specific environment requires model training, the proposed method can be applied without such preparation. From the above, the proposed method is considered a more efficient method than the related works' methods.

Secondly, some parameters had to be adjusted according to the background. The thresholds L_0 and L_1 to extract frames from the video had to be changed depending on the complexity of the background, as shown in Fig. 14. Since the frame with fluctuation through L_0 and L_1 is extracted, if the setting for the flat background in Fig. 14 (2) is applied to the complicated background in Fig. 14 (1), the number of extracted frames decreases. Its reverse is also the same. Therefore, in this experiment, I created a tool to monitor the transition of the difference shown in Fig. 14 and set the threshold manually.

Similarly, as mentioned in Sec. 5.2, as for the threshold for the binarization shown in Figs. 12 and 13, higher accuracy was obtained by setting a small threshold of 159 than a large one of 191 for the flat background. It is expected that the above parameters can be set automatically by analyzing the transition of the difference in Fig. 14. This is a future challenge.

7 CONCLUSIONS

We can efficiently collect data for training and discrimination for deep learning using mobile camera videos such as wearable cameras. For a small target, it is necessary to extract a relatively small area that includes the target from the video frame, since the background's influence decreases the discrimination accuracy. However, it was often difficult to extract a moving target area efficiently from a video shot by a moving camera in a complex environment.

Hence, I proposed a method for extracting the target area by creating a subtraction of target images between adjacent frames. In this method, the backgrounds are aligned based on the displacement vector in the optical flow and superimposed. Moreover, through the experiment, I showed that a certain accuracy could be achieved even in complex backgrounds. Furthermore, taking advantage of consecutively extracting the target area from the video, I also showed the accuracy could be improved by comparing the extracted area size with its moving average.

Future studies will focus on applying this method to image recognition using deep learning and verifying its effectiveness.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 19K11985.

REFERENCES

[1] F. Chollet, "Deep learning with python," Manning Publications Co. (2017).

[2] A. Elqursh, and A. Elgammal, "Online moving camera background subtraction," European Conference on Computer Vision, pp. 228–241 (2012).

[3] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," Scandinavian conference on Image analysis, Springer, pp. 363–370 (2003).

[4] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, "Deep learning," MIT press (2016).

[5] G. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," Proc. IEEE Int. Conf. on computer vision workshops, pp. 142–150 (2015).

[6] M. Jain, H. Jegou, and P. Bouthemy, "Better exploiting motion for better action recognition," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2555–2562 (2013).

[7] A. Kaehler, and G. Bradski, "Learning OpenCV 3: computer vision in C++ with the OpenCV library," O'Reilly Media, Inc. (2016).

[8] A. I. Károly, R. N. Elek, T. Haidegger, K. Széll, and P. Galambos, "Optical flow-based segmentation of moving objects for mobile robot navigation using pre-trained deep learning models," 2019 IEEE Int. Conf. on Systems, Man and Cybernetics, pp. 3080–3086 (2019).

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097–1105 (2012).

[10] T. Kudo, and R. Takimoto, "CG utilization for creation of regression model training data in deep learning," Procedia Computer Science, Vol.159, pp. 832–841 (2019).

[11] T. Kudo, "A proposal for article management method using wearable camera," Procedia Computer Science, Vol. 176, pp. 1338–1347 (2020).

[12] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017).

[13] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection." Proc. IEEE Int.l Conf. on Computer Vision, pp. 2980–2988.(2017).

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector,".European Conf. on Computer Vision, pp. 21–37, Springer. (2016).

[15] K. Makino, et al., "Moving-object detection method for moving cameras by merging background subtraction and optical flow methods," 2017 IEEE Global Conf. on Signal and Information Processing. pp. 383-387 (2017).

[16] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for IoT big data and streaming analytics: A survey," IEEE Communications Surveys & Tutorials, Vol. 20, No. 4, pp. 2923–2960 (2018).

[17] M. Narayana, A. Hanson, and E. Learned-Miller, "Coherent motion segmentation in moving camera videos using optical flow orientations," Proc. IEEE Int. Conf. on Computer Vision, pp. 1577–1584 (2013).

[18] OpenCV team, "OpenCV modules," https://docs.opencv.org/4.3.0/index.html (referred May 25, 2020).

[19] J. Redmon, S.Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 779–788.(2016).

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," Advances in Neural Information Processing Systems, pp. 91–99 (2015).

[21] S. Singh, C. Arora, and C. V. Jawahar, "Trajectory aligned features for first person action recognition," Pattern Recognition, Vol. 62, 45-55 (2017).

[22] M. Unger, M. Asbach, and P. Hosten, "Enhanced background subtraction using global motion compensation and mosaicking," 2008 15th IEEE Int. Conf. on Image Processing, pp. 2708–2711 (2008).

[23] M. Verhelst, and B. Moons, B., "Embedded deep neural network processing: Algorithmic and processor techniques bring deep learning to iot and edge devices," IEEE Solid-State Circuits Magazine Vol.9, No. 4, pp. 55–65 (2017).

[24] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proc. 2001 IEEE computer society conference on computer vision and pattern recognition, Vol. 1, pp.511-518 (2001).

[25] H. Wang, H, and C. Schmid, "Action recognition with improved trajectories," Proc. IEEE Int. Conf. on Computer Vision, pp. 3551–3558 (2013).

[26] M. Yazdi, and T. Bouwmans, "New trends on moving object detection in video images captured by a moving camera: A survey," Computer Science Review, Vol. 28, pp. 157–177 (2018).

[27] W. Zhang, X. Sun, and Q. Yu, "Moving object detection under a moving camera via background orientation reconstruction," Sensors, Vol. 20, Issue 11, 3103 (2020).

[28] Q. Zhao, T. Sheng, Y. Wang, Z. Tang, Y. Chen, L. Cai, and H. Ling, "M2det: A single-shot object de-

tector based on multi-level feature pyramid network," Proc. AAAI Conf. on Artificial Intelligence, Vol. 33, pp. 9259–9266.(2019).

(Received October 16, 2020)
(Accepted May 11, 2021)

Tsukasa Kudo received the BSc and ME from Hokkaido University in 1978 and 1980, and the Dr. Eng. from Shizuoka University in 2008. In 1980, he joined Mitsubishi Electric Corp. He was a researcher of parallel computer architecture and engineer of business information systems. Since 2010, he is a professor of Shizuoka Institute of Science and Technology. Now, his research interests include deep learning and database application. He is a member of IEIEC and IPSJ.

<u>**Industrial Paper**</u>

High-performance Synchronized Control between Spindle and Servo Motors for CNC Equipment

Kazuhiko Tsutsui[*†], Katsuhiko Kaji[††], Katsuhiro Naito[††],
Naoya Chujo[††] and Tadanori Mizuno[††]

[*]Mitsubishi Electric Corporation
[†]Graduate School, Aichi Institute of technology, Japan
[††] Aichi Institute of technology, Japan
[*]Tsutsui.Kazuhiko@ay.MitsubishiElectric.co.jp,

Abstract -
In recent CNC machine tools, it is necessary to control the spindle and the servo axis in synchronization with each other, for example thread cutting or synchronous tapping, which has become more important.

Conventionally, position and speed loops are generally controlled on the CNC side, and the configuration of the master-slave follow-up control is adopted in which the servo motor follows the less responsive spindle motor. However, this method has three limitations: the command response due to the feedback of the spindle motor, the characteristic variation of the Induction Motor (IM motor), and the influence of the network dead time.

In this paper, we propose the high-performance synchronized control method between spindle and servo. This method consists of mainly three components. First, to improve the characteristic of IM motor control, we develop a unique multi-core system which minimize the dead time in the position/speed/current control loop for the spindle motor. Second, we propose a control method to improve the characteristics of IM motor, whose characteristics tend to change with temperature. Finally, we propose a compensation method using a high-speed network to minimize synchronization errors caused by differences in responsiveness.

By the proposed method, we achieved much higher productivity of machining process especially for the synchronized cutting process such as Thread cutting and Synchronous Tapping. Compared with the conventional control, productivity was improved by 20%. Besides, the system became robust to power supply environment.

1 INTRODUCTION

The CNC system is roughly composed of four components. The first is a CNC controller that configures machine coordinates according to the machining program generated by the user and generates the position and speed commands for the feed axis and spindle (tool). The second is actuators such as servo motors mainly for position control of table on which work is mounted and spindle head on which rotating tool is mounted. And one of the other actuators is the spindle motor mainly for speed control by rotating tool at high speed in the machining center, or rotating a workpiece itself in the lathe. Third is the servo amplifiers and spindle amplifier which are power converters that supply power to the actuator. Finally, there are detectors such as encoders and linear scales that feedback the position and speed of the operating part of the machine or motor.

The brain for the whole control of the machine tool is a CNC. This controller mainly generates the movement amount of each feed axis per unit time by analyzing sequentially the machining programs called G-code which describes the machining path of the tool, the feed speed of the tool and the number of revolutions of the tool. Further, a command value of the rotation speed of the spindle motor is generated according to the cutting conditions.

Here, the spindle motor must perform a milling process that requires an ultra-high speed of over 30,000 revolutions per minute. In addition, high power exceeding several tens of kilowatts to withstand heavy cutting is required. Therefore, the induction motor (IM motor) that does not use the permanent magnets is used for the spindle motor, instead of the synchronous motor (SM motor) with the permanent magnets that is generally used for a servo motor.

Therefore, in the spindle motor control, it is difficult to realize high-response control at the same level as the servo. This means that it is difficult to completely synchronize the rotational position of the spindle and the position of the servo feed, such as thread cutting and synchronous tapping.

In the conventional threading process, CNC generates the feed axis position command as a master based on the rotational position of the spindle that response slowly and a relatively responsive servo axis follows the position of the slow spindle. In this method, it is necessary to limit that the processing speed is slow and acceleration is small.

Further, in the case of synchronous tapping, since the cutting load is relatively small and the inertia of the cutting tool (Tapper) itself is small, unlike in the case of thread cutting in which the workpiece itself is rotated, position commands to the spindle and the servo are given independently in some cases. However, even in this case, if there is a slight difference in synchronization due to the effect of the biting of the facet, the tapper may be damaged, and a screw conforming to the standard cannot be processed.

In this paper, section 3.1 of chapter 3 proposes a method for constructing a high-response, high-precision feedback loop that increases the responsiveness of the servo and the spindle itself and prevents the fluctuating in position and speed due to high-speed machining and cutting disturbances.

The most important point in constructing a high response feedback loop is minimizing the dead time and the processing cycle time in the loop.

Therefore, we have implemented position, speed, and current loop control in servo amplifiers and spindle amplifiers in order to minimize the effects of dead time caused by the communication interface outside feedback loop control.

In addition, in order to minimize the time delay and processing time in the servo and the spindle amplifier and to implement the function to improve the motor control performance in a specialized method, a multi-core design was proposed, and the feasibility of high-speed and high-precision control of the spindle and the servo was verified in section 4.1.

Next, in section 3.2, we proposed a control method that improves the characteristics of the IM motor, whose characteristics tend to change in the temperature environment, and stably achieves the performance of the spindle motor, and verified its effectiveness in section 4.2.

Finally, section 3.3 proposes the method to realize compensation between amplifiers to compensate for the difference in response between the servo and the spindle that occurs even when the response of the servo and spindle is increased. As for the effect of this, section 4.3 shows the verification results of this compensation. Subsection 4.3.1 describes in threading, and Subsection 4.3.2 and 4.3.3 show the verification results in actual synchronous tapping.

2 RELATED TECHNOLOGIES

Chapter 2 shows the CNC system and the basic control structure of the servo and spindle, and then describes the problems of conventional spindle and servo synchronous control.

2.1 Basic Configuration of CNC Equipment

Multi-tasking machines such as the one shown in Fig. 1

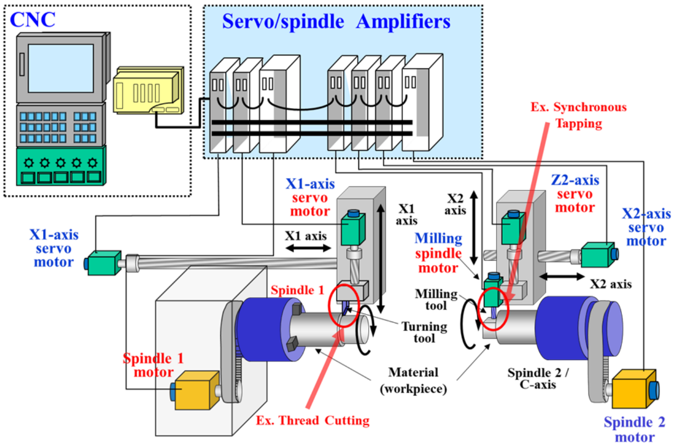


Figure 1: System configuration example of multi-tasking machine

are increasing in recent years.

Each axis is driven by servo motor connected to the ball screw in the machine. The tool used for cutting is attached to the spindle head and driven by the spindle motor. The CNC controller and the amplifiers that are used to control the speed and position in addition to power supply for the servo and the spindle motors are all installed in an electrical enclosure.

In the CNC system, since the path of the cutting tool directly affects the accuracy of the workpiece, it is important to follow the command with a small error against the influence of various load disturbances such as cutting disturbance and machine friction.

It is also important to match the performance of synchronization and response between the X, Y, and Z axes. If the synchronization and response of each axis do not match, the tool path is not able to follow the command from CNC controller. It means that the processing (cutting) accuracy could not satisfy the required quality.

2.2 Basic Servo/Spindle Control Architecture (Distributed Control)

Figure 2 shows a control block diagram at the time of thread cutting in which the rotational position of the main spindle and the position of the servo axes need to be synchronized.

Permanent magnets are used for servo motors that require high-response position control, while induction motors that do not use permanent magnets are generally used for spindle motors. For example, in a machining center or lathe that uses CNC, it is possible to rotate a rotating tool more stably higher speed than a case that requires position control of a spindle motor, or to withstand heavy cutting well. Since it is important to achieve the highest possible output, synchronous motors that are not suitable for high-speed rotation or high-output at high-speed rotation are rarely used.

On the other hand, machining required for machine tools

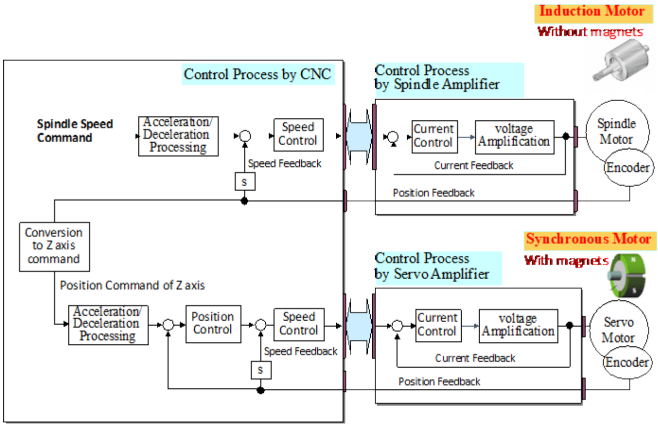


Figure 2: Synchronous control between Spindle and Servo（Thread Cutting）

is becoming more complicated, and machining that requires synchronizing the positions of the spindle motor and the servo motor, such as threading and synchronization taps, must be performed by one machine.

As described above, the response of the spindle motor is low which means that the response frequency is one digit lower than servo motors. Therefore, the method of centralized control as shown in Fig. 2 is common in general. The servo axis uses the position feedback of the spindle motor which is returned to the CNC side as a position command of the servo axis in such thread cutting.

As described above, the servo follows the movement of the spindle having low response, so that the synchronization error between the spindle and the servo can be reduced.

2.3 Issues with Spindle and Servo Synchronous control

In this section, the problems in the conventional synchronous control between spindle and servo are shown by using the example of thread cutting control and synchronous tap machining.

2.3.1 Thread Cutting Control

It is possible to improve the synchronization performance between the spindle and the servo axis by constructing a control method as shown in Fig. 2 and making the servo with relatively quick response follow the position of the spindle motor with low response. Since a spindle with low response is used as a reference, there is a problem that unless the response of the spindle increases, the tact time cannot be improved by increasing the speed of thread cutting.

In addition, the position feedback of the spindle motor is returned to the CNC side once, converted to the servo axis command, and then passed again to the servo amplifier via the network, which increases waste time. Therefore, there was a limit to suppresses synchronization errors eventually.

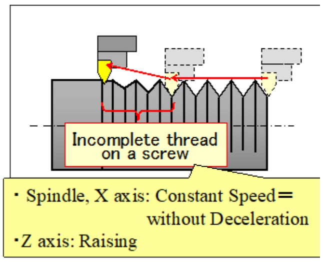


Figure 3: Incomplete thread on a screw

Further, since there is a difference in the acceleration/deceleration time between the spindle and the servo, machining cannot be started until the spindle and the servo reach a constant speed in order to ensure synchronization accuracy. Alternatively, there is also a problem that the speed of the main spindle and the X-axis cannot be reduced until the Z-axis finishes to raise as shown in Fig. 3.

2.3.2 Synchronous Tapping

Also, in synchronous tapping, similar to thread cutting, since it is necessary to control the servo axes in synchronization with the rotational position of the main spindle, centralized control on the CNC side as shown in Fig. 2 is adopted in some cases. On the other hand, compared to the thread cutting, since the drilling is performed after the preliminary hole is formed in advance, the cutting load is relatively small, and unlike the thread cutting which rotates the work itself, the inertia of the cutting tool (tapper) itself is small.

In some cases, as shown in Fig. 4, position commands to the spindle and the servo are given independently of each other, and in many cases, the configuration is aimed at higher-speed machining.

However, if there is a slight deviation in synchronism such as the influence of the bite of a facet, there are problems such as inducing damage to the tapper and making it impossible to process a screw conforming to the standard.

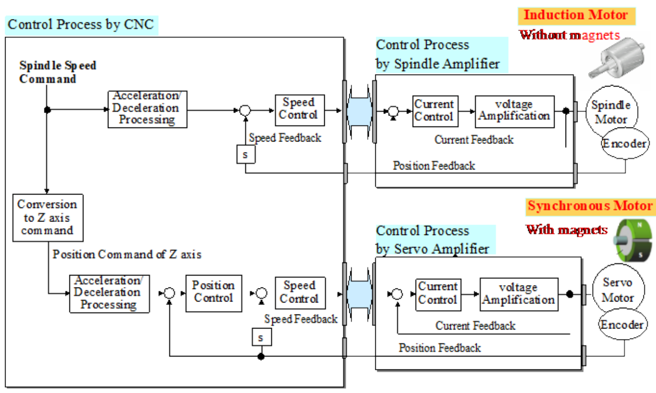


Figure 4: Synchronous control between Spindle and Servo（Synchronous Tapping）

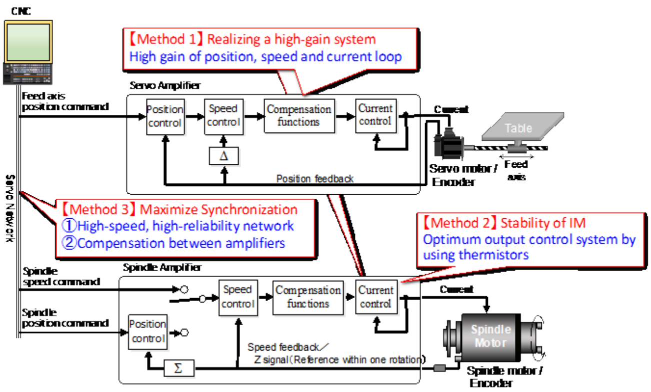


Figure 5: Basic configuration of High-response synchronous control between Spindle and Servo

3 PROPOSED SYSTEM (High-response synchronous control system between Spindle and Servo)

This section describes the proposed methods to realize the synchronous control between the spindle and servo with high response and accuracy. The system is largely composed of the following three methods, the basic structure of which is shown in Fig. 5.

As mentioned above, in the conventional system, there have been many cases in which the controller controls the position and speed loop and switches the control method according to the machining process, especially in the threading and synchronous tapping in which synchronous control between the spindle and servo is required. In this paper, in order to pursue the high-speed command follow-up characteristics and the robustness against disturbance, we propose the distributed control system in which the position and speed loop are executed by the servo and spindle amplifier.

Based on this decentralized control, we could achieve the high-precision synchronous control system between the spindle and the servo by using the following three methods.

[Method 1] High gain of position, speed and current loop

Section 3.1 describes the details of Method 1 for realizing a high-gain system in which each independent spindle and servo follow commands from the controller at high speed and with high response and are not easily affected by cutting disturbances.

[Method 2] Power optimization of IM spindle motor using thermistors

Induction motors (IM motor) used in the spindle are greatly affected by temperature characteristics. That is because it is generated by the motor's own coil and magnetic circuit, which is different from a synchronous motor in which the magnetic field for generating torque can be obtained from a stable permanent magnet.

Therefore, we have developed an optimum output control system using a thermistor, and its effect is described in Section 3.2.

[Method 3] High-speed, high-reliability network and compensation between amplifiers

Section 3.3 describes compensation between amplifiers to maximize synchronization performance by complementing the difference in responsiveness between the spindle and servo, and also describes high-speed, high-reliability networks to realize this.

3.1 PROPOSED METHOD 1 (High gain of position, speed and current loop)

The following items are necessary to achieve high gain performance in the main spindle and servo control loop.

(1) Faster position, speed, and current loop processing
(2) Reduce the dead time in the loops
(3) Higher accuracy and higher resolution of position, speed and current feedback data

In practice, we have implemented all of the initiatives in (1), (2), and (3) to realize the system. In this paper, we describe the initiatives in (1) that are particularly distinctive.

Figure 6 shows a block diagram of the newly developed the current control core unit. Conventionally, the processing of the position, speed, and current loop in the spindle and servo amplifier has been sequentially performed by software processing by a general-purpose CPU.

On the other hand, in order to greatly improve the performance, we have developed a dedicated calculation unit which is called the current control core unit that is the innermost among the control loops and requires the highest gain (high response).

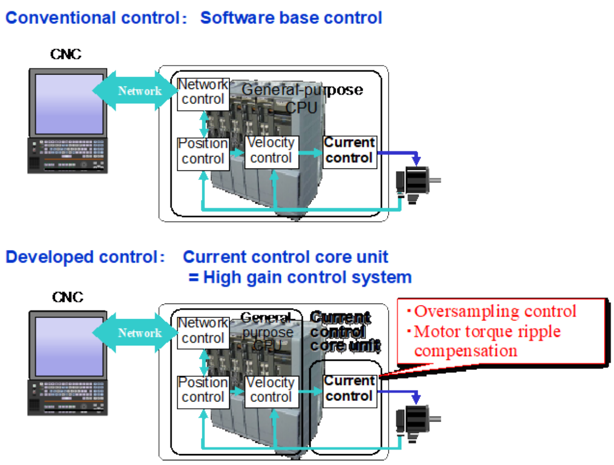


Figure 6: Current control core unit

Thus, oversampling function enables parallel running of position/speed loop processing and current loop processing, so that the PWM switching frequency can be stably increased, and high gain of the position/speed/current loop can be realized.

3.1.1 Oversampling Function

Figure 7 shows a timing chart of the oversampling function mounted on the current control core unit. As a comparison, the timing of the valley sampling which was performed by the conventional control is described. In the conventional control, the position control, speed control and other various processes are performed by one CPU, so that the current loop cannot be always operated. Therefore, the current feedback (current FB) data is AD converted only at the valley timing of the reference triangular wave, which is least affected by the PWM switching. On the other hand, by processing the current loop control in the current control core unit, the current loop can be constantly turned, and at the same time, the ΣΔ + IIR filter is incorporated in the current control core unit instead of the one-time conversion by the AD converter, so that the current can be detected with less ripple and with high accuracy.

Here, we adopted that the resolution of ΣΔ + IIR filter is corresponding with 12bits, the carrier frequency is 9kHz and 5 times oversampling.

3.1.2 Motor Torque Ripple Compensation

Figure 8 is a control block diagram of the current control core unit. As shown in the figure, the compensation of harmonic wave is incorporated in the dq-axis via UVW phase conversion on the current feedback side and the voltage command side. The torque generated by the motor is generated by the product of the linkage magnetic flux of the rotor and stator and the motor current. At this time, the torque ripple is generated by the interlinkage magnetic flux containing many harmonic components. For this reason, a

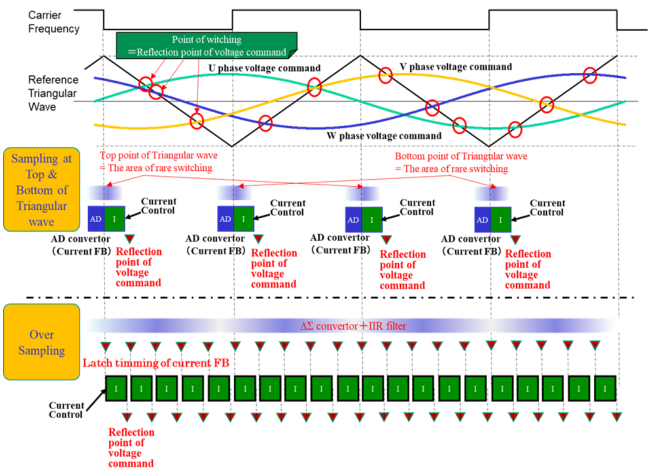


Figure 7: Oversampling for current loop

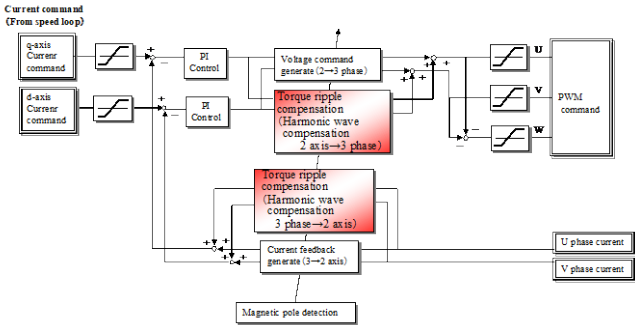


Figure 8: Torque ripple compensation in the current control core unit

function to correct this problem is also built in for the current control core unit.

3.2 PROPOSED METHOD 2 (Output optimization of IM main Spindle motor utilizing thermistor)

The torque τ of an induction motor (IM motor) generally used for a spindle motor is given by the following equation.

$$\tau = \frac{PM}{2L} i_q \psi_d$$

$$\psi_d = M i_d$$

τ : Torque of the induction motor

P: Number of poles

M: Mutual inductance between the stator and rotor

L: Inductance

ψ_d: interlinkage flux

i_q: q-axis current, i_d: d-axis current

Here, since the interlinkage magnetic flux is made of the product of the d-axis current flowing in the motor and the inductance instead of a permanent magnet like a servo motor (Synchronous motor), it is easy to be affected by temperature, and it is difficult to always obtain a stable output.

Therefore, as shown in Fig. 9, we developed the system that a thermistor is built into the coil of the spindle motor, and it enables the spindle amplifier to constantly monitor

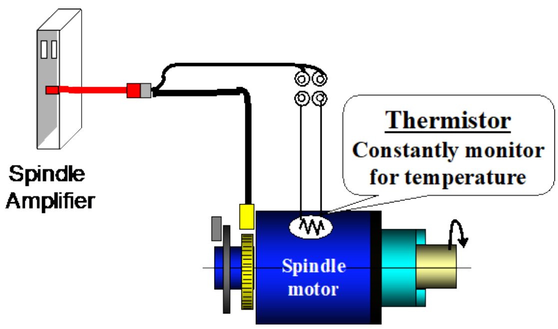


Figure 9: Constant monitor for temperature

the temperature. In this system, stable output characteristics are obtained regardless of temperature by controlling the d-axis current i_d according to temperature conditions.

3.3 PROPOSED METHOD 3 (High-speed, high-reliability network and compensation between amplifiers)

In order to maximize the advantage that the servo and spindle control gain increases, such as the tracking performance in high-speed cutting, it is necessary to improve the cycle of the command output from the CNC controller to the spindle and servo amplifiers. Besides, it is also important to increase the accuracy of command units.

Another advantage of using the distributed control method is that the dead time can be minimized because the network is not interposed in the control loop of the servo or spindle. On the other hand, there is a problem that compensation between the servo amplifiers or between the spindle and the servo cannot be performed.

In this section, we have developed a high-speed, high-reliability network and the compensation between amplifiers.

3.3.1 High-speed, High-reliability Network

We have developed a high-speed optical communication servo network that dramatically improves the network performance between the CNC-servo and the spindle.

The points of this network are as follows.

- Communication baud rate improvement:
  5.6 MHz → 50 MHz (Approximately 10 times)
- Improvement of communication cycle:
  1.7 msec → 0.2 msec (about eight times)
- Improvement of command resolution: 1 μm → 1 nm
- Protocol that enables data exchange between servos and spindle

3.3.2 Improvement of Synchronous Control between Spindle and Servo by the Compensation between Amplifiers

In section 3.1, we describe how to improve command followability and rigidity against external disturbances by realizing high gain control of each of the servo and spindle.

However, in the spindle control, there are cases where the command follow-up characteristics are inferior to the servo because an induction motor (IM motor) which is difficult to achieve high responsiveness due to the influence of an electrical time constant and high inertia is used and the acceleration/deceleration characteristics are not stable due to voltage saturation (torque saturation) caused by a power source environment or the like. It means that it is sometimes difficult to obtain high synchronization performance between the spindle and the servo axis.

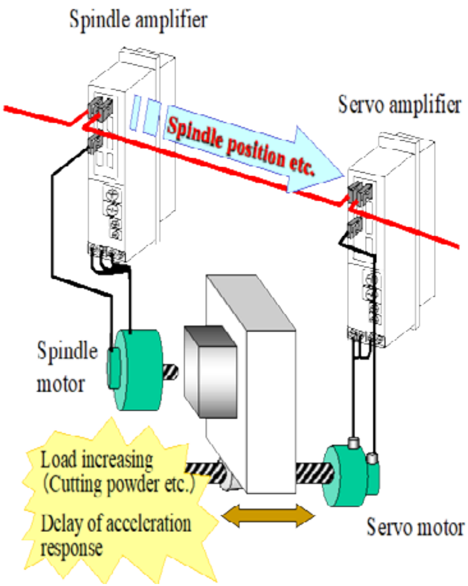


Figure 10: High-speed synchronous tapping function

Therefore, we have developed a high-speed synchronous tap function that utilizes the data communication protocol between amplifiers incorporated in a high-speed optical network and utilizes the amplifier compensation between the spindle and servo as shown in Fig. 10.

Figure 11 shows a block diagram of the high-speed synchronous tap function. A position command synchronized with the spindle and the servo is sent via a network. On the other hand, the spindle amplifier and the servo amplifier perform position loop control to control the motor to follow the command.

If it is possible to ideally follow the command without any cutting disturbance or the like, synchronization accuracy can be guaranteed, but in general, it is difficult to improve the speed frequency response of the spindle motor control more than that of the servo control shaft.

Therefore, in the proposed method, the position deviation and the speed feedback in the spindle amplifier are passed to the servo amplifier using the inter-amplifier data reception protocol provided in the high-speed optical network. On the servo amplifier side, unit conversion is performed using the spindle position sent by the CNC controller and the servo axis conversion coefficient K which means a ball screw pitch, and the spindle position deviation is added to the servo axis position command as a correction position, and the spindle speed feedback is added to the speed command as a correction value. The spindle speed feedback is further differentiated, converted by the inertia J of the servo and the torque constant of the motor, and then added to the current command as a correction current value. Since these correction position commands and current commands are data delayed by the network, they are values fed forward to make up for the dead time Td caused by the network.

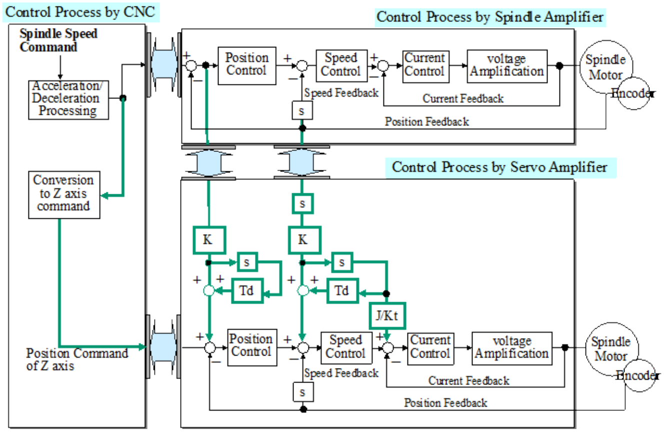


Figure 11: Diagram of High-speed synchronous tapping

As a result, it is possible to achieve high precision and high speed in the servo-to-spindle synchronous control such as the synchronous tapping compare with the conventional control, and this can greatly contribute to the improvement of productivity.

4 EVALUATION

First, the effects of method 1 and 2 which are measures to increase the response of each spindle and servo are shown in Section 4.1 and 4.2. Then, the synchronous control performance of the spindle and servo with high response including the effects of method 3 is shown based on the simulation results of thread cutting and the actual machining results of synchronous tapping in Section 4.3.

4.1 Verification of Effects of High gain system (Method 1)

First, the effect of increasing the control loop gain of the spindle and servo using the current control core unit equipped with an oversampling function and torque ripple compensation is shown.

Figure 12 shows the result of temperature rise of the motor and the result of frequency spectrum analysis of the motor current when the spindle motor is rotated at high speed of 12,000 r/min. The harmonic component superimposed on the current is drastically cut. As a result, the loss called iron loss generated in the motor is reduced. As a result, the temperature is reduced by 10%.

Figure 13 shows the result of roundness measurement in the X-Y plane of a servo-driven machine tool. Under the condition that the accuracy of the circle was 4.0 μm in the conventional control, the accuracy of the circle was improved 1.5 times to 2.5 μm by increasing the gain through the method implemented this time.

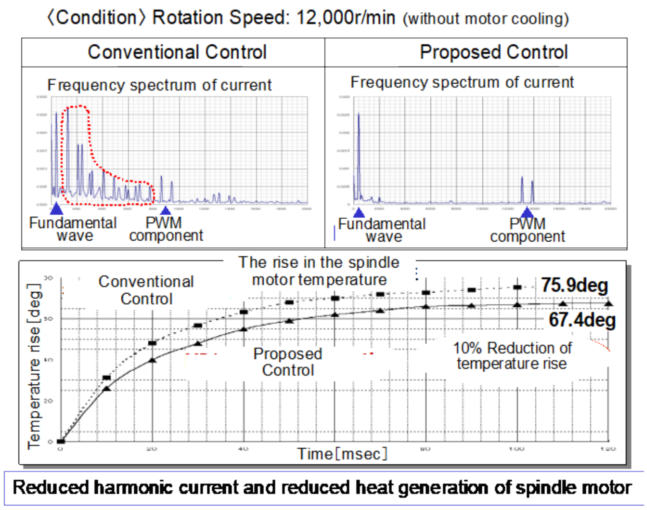


Figure 12: The effect of High-gain control (Spindle)

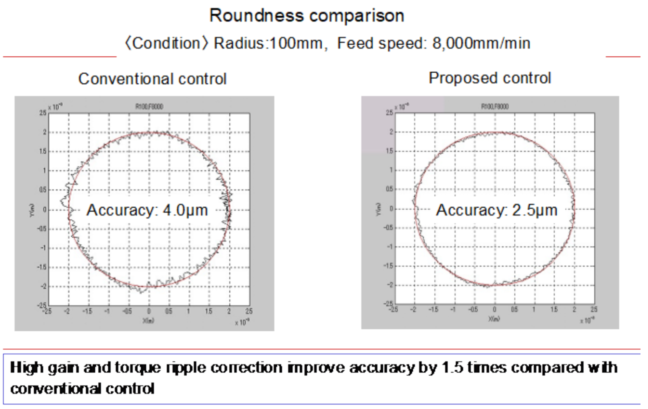


Figure 13: The effect of High-gain control (Servo)

4.2 Effect of Output Optimization Control of IM Spindle Motor (Method 2)

Figure 14 shows the effect of adopting a new control system that controls output characteristics by utilizing information from a thermistor mounted on the spindle motor. The plot on the right shows the acceleration /deceleration

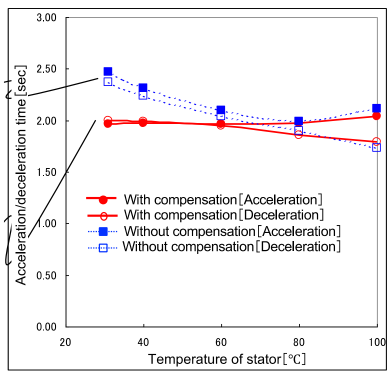


Figure 14: The effect of the Spindle temperature compensation based on thermistor

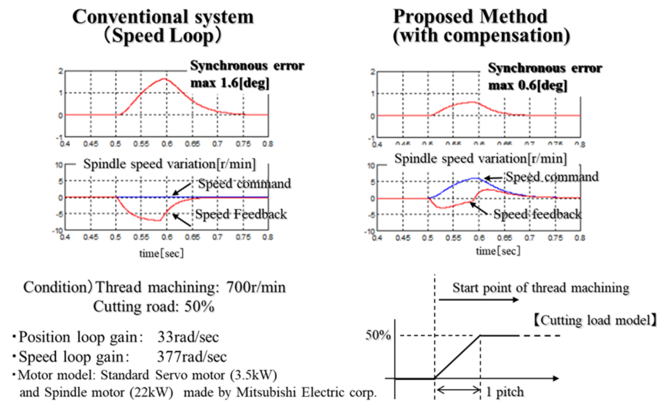


Figure 15: Effect in case of thread machining

time change when the temperature changes after adjusting the parameter under the condition of 80 ℃. The blue dotted line plots the fluctuation during the conventional control without compensation and the red solid line plots the fluctuation when the proposed compensation is performed. The figure shows that the fluctuation due to the temperature change is suppressed.

4.3 Verification of Synchronous control Performance between Spindle and Servo

For the performance improvement of synchronous control, subsection 4.3.1 describes the evaluation by simulation, and subsection 4.3.2 describes the evaluation by real machine.

4.3.1 Thread Machining Improvements: (Simulation Validation)

In this subsection, we simulated the thread machining improvements by using Matlab/Simulink before real machining (cutting) process. Because it is easy to assess the advantage of proposal method by simulation instead of real machining process.

Figure 15 shows the simulation results of threading by synchronous control of the spindle and servo. The simulation condition is follows.

[Simulator] Matlab/Simulink

[Control loop gain]

Position loop gain: 33rad/sec

Speed loop gain: 377rad/sec

[Motor model]

Standard servo motor (3.5kW) and spindle.

Motor (22kw) made by Mitsubishi Electric corp.

[Machining condition]

Speed: 700r/min, Cutting load: 50%

This time, the conventional system in which the servo follows the rotational position feedback of the spindle has been changed to a system in which the spindle and the servo

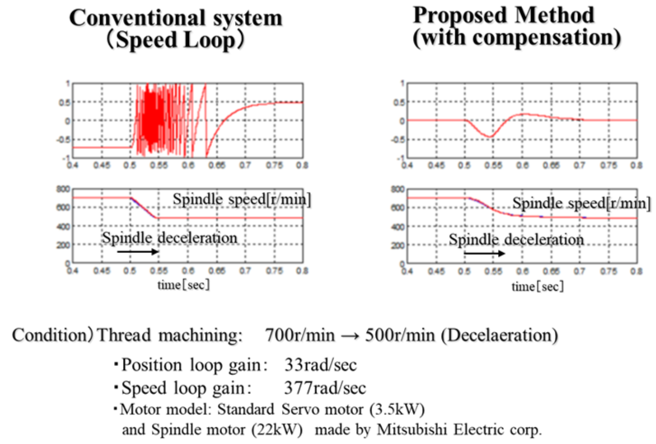


Figure 16: Synchronous accuracy in case of thread machining (Overriding speed conditions)

follow the position command. Synchronization accuracy is improved by performing compensation between the spindle and servo in conjunction with high gain control. The synchronous error became 0.6 degree from 1.6 degree which was the result of conventional method.

As a result, conventionally, in order to ensure synchronization accuracy, machining cannot be started until the spindle and servo reach a constant speed because synchronous error is too much to use, but in the proposed method, the synchronization accuracy can be improved within appropriate level (0.6 degree) even when the spindle speed is changed as shown in Fig. 16 (Overriding speed condition). Here, Fig. 16 case shows the condition when the machining speed was decreased from 700r/min to 500r/min.

As shown in Fig. 17, the spindle can be decelerated (Override) when the Z-axis tool is pulled up, and the length of the incomplete threaded portion can be shortened.

Figure 17 shows the calculation result the condition of which is follows.

[Machining model]

M56 screw (pitch: 5.5mm, depth of cut: 4mm)

[Spindle speed] 1,000r/min

[X axis speed] 5.5m/min, [Ball screw pitch] 10mm

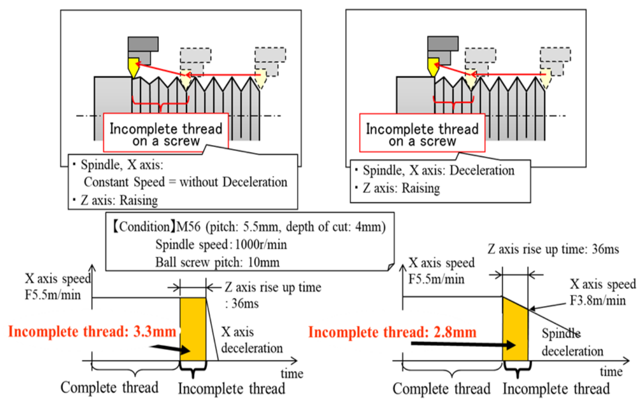


Figure 17: Reduction of incomplete thread

4.3.2 Verification of Synchronous Control Performance between Spindle and Servo (High-speed and high-precision synchronous tapping: Verification of real machine machining)

Figure 18 shows the results of synchronous tapping in an actual machining center. The figure on the left shows the results under the conventional operating conditions in which the compensation function between amplifiers is disabled. Here, since the spindle motor needs to be used in a region where torque saturation does not occur, the motor is operated with an acceleration/deceleration time constant of 560 msec including a margin. On the other hand, the middle figure shows a case in msec. At this time, the torque saturation region of the spindle motor is applied, so that the spindle motor cannot follow the command and the synchronization error with the servo axis increases from 30 pulses to 160 pulses.

On the other hand, the figure on the right shows the result when the compensation between amplifiers is enabled. As shown in the middle figure, the time constant is 350 msec, which indicates that the torque of the spindle motor is saturated. However, the figure shows that the synchronization error between the spindle motor and the servo axis is greatly improved to 8 pulses.

Figure 19 shows a plot of the synchronization error when the tapping of the M6 screw is cut by changing the time constant. In the normal control, the synchronization error increases as the acceleration/deceleration time constant of the spindle decreases, and especially when the time constant is set so as to be small as to enter the torque saturation region, the increase of the synchronization error becomes remarkable. In other words, in actual use, it is necessary to set the time constant with enough margin so that torque saturation does not occur in consideration of changes in the power supply environment. On the other hand, when the compensation between amplifiers is enabled, the synchronization error is suppressed even if it is applied to the torque saturation region. It means that it is not necessary to set the time constant with enough margin as in the past, and it can greatly contribute to shortening the machining time.

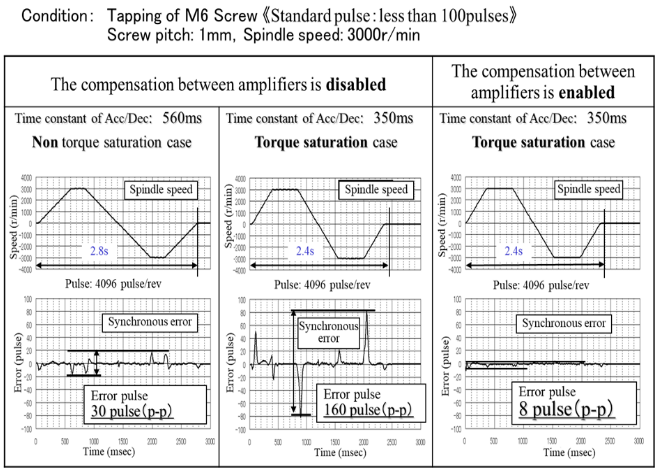


Figure 18: Verification result of synchronous tapping

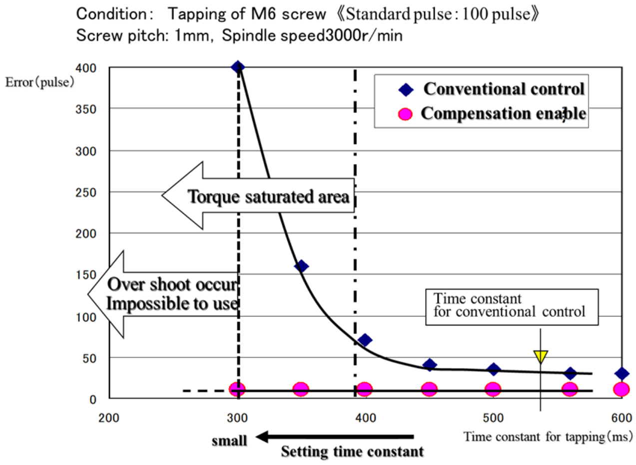


Figure 19: Time constant for tapping and synchronous error

4.3.3 Verification of Synchronous Control Performance between Spindle and Servo (Continuous machining of synchronous tapping: Verification of actual machine machining)

Figure 20 shows the results of continuous machining of synchronous tapping of M5 screws on an actual machining center. Even if the spindle speed and acceleration /deceleration time constant are shortened while securing the screw accuracy under the conventionally set machining conditions without the compensation between amplifiers, the accuracy can be secured by performing the compensation between amplifiers, resulting in productivity improvements of 30 to 36 units and 20% within the same time (75 seconds).

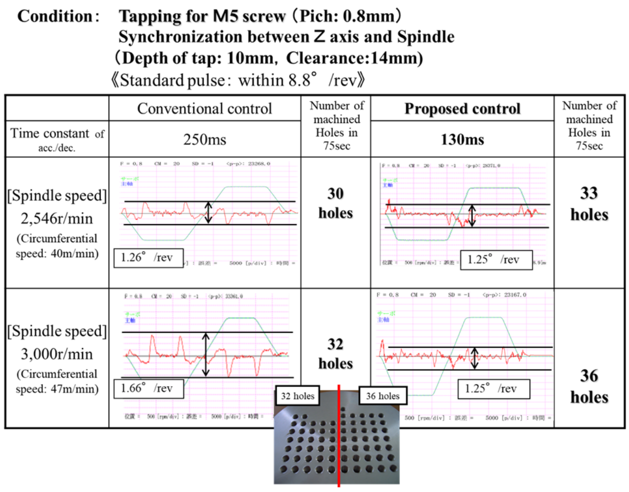


Figure 20: Result of continuous tapping by proposed control

5 CONCLUSION

In recent years, in order to improve productivity, the demand for combined machining corresponding to various machining operations by one machine tool has been increased. Under these circumstances, there has been an increase in the number of cases that the spindle and the servo are used for machining in synchronization. This paper introduces the methods implemented to realize high speed and high precision in the synchronous machining of the spindle and servo.

At first, we proposed a method for constructing a high-response, high-precision feedback loop that increases the responsiveness of the servo and the spindle itself and prevents the fluctuating in position and speed due to high-speed machining and cutting disturbances.

Specifically, in order to minimize the time delay and processing time in the servo and the spindle amplifier and to implement the function to improve the motor control performance in the specialized methods, a multi-core design was proposed, and the feasibility of high-speed and high-precision control of the spindle and the servo was verified by showing below. In the case of the spindle motor control, the effect of reducing heat generation of the spindle motor for suppressing deterioration in accuracy due to thermal expansion of the machine is shown. In the case of the servo control, the effect of improving the accuracy of the feed shaft is shown.

Next, we proposed a control method that improves the characteristics of the IM motor, whose characteristics tend to change in the temperature environment, and stably achieves the performance of the spindle motor, and verified its effectiveness for the fluctuation suppression of acceleration and deceleration time in a real machine.

Finally, we proposed the method to realize compensation using the high-speed and high-reliability network between amplifiers to compensate for the difference in response between the servo and the spindle that occurs even when the response of the servo and spindle is increased. As for the effect of this, we showed the simulation verification results in threading, and showed the verification results in actual synchronous tapping.

As a result, in comparison with the conventional control, for example, under the conditions of the synchronous tap processing shown in this document, productivity improvement of 20% was achieved. Besides, it is able to make a robust system to power supply environment.

This also served as a guideline for making the most of the future evolution of semiconductor processes for spindle and servo control, and at the same time showed the importance of the network between the CNC controller and spindle and servo.

In the future, we plan to continue to optimize the architecture and compensation algorithms to maximize the synchronization performance of the spindle and servo, while also incorporating the evolution of semiconductors and network technologies timely.

REFERENCES

[1] T. Tanaka, K. Tsutsui: "Driving system MDS-D/DH series for M700", Mitsubishi Electric Advance, Vol.116, pp. 23-25, (2006).

[2] H. Sugimoto, M. Koyama, S. Tamai: "The theory of AC servo system and real design", General electronic publisher, p.31-71 (1997).

[3] S. Furutani, A. Satake: "Proposal of current control for high speed AC motor control", Journal of the Institute of Electrical Engineers, D, Vol.128, No.12, p.1361-1402, (2008).

[4] S. Furutani, A. Satake: "Current control method of AC motor for low carrier frequency drive", Journal of Lectures at the National Congress of the Institute of Electrical Engineers, p.224-228, (2006).

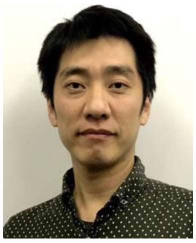
[5] R. Toutant, S. Balakrishnan, S. Onyshko, and N. Popplewell: "Feedrate Compensation for Constant Cutting Force Turning", IEEE Control Systems Magazine, Vol.13 , No. 6 , (1993).

[6] D. Kim, C. Hyuk Yim : "All digital high performance controller for spindle motor in CNCmachine tool", IEEE International Electric Machines and Drives Conference Record, pp. MC2-2.1-MC2-2.3, (1997)

[7] Y. Nakamura, S. Futami: "Application of predictive control for FA mechatronics equipment", Journal of the Society of Instrument and Control Engineers, Vol.39, No.5, (2005).

(Received October 18, 2020)
(Accepted July 28, 2021)

Kazuhiko Tsutsui received the B.E. and M.E. degrees in Mechanical Engineering from Kumamoto University, Japan in 1990 and 1992 respectively.
In 1992, he joined Mitsubishi Electric Corp. Since when he joined Mitsubishi Electric Corp, he has been engaged in development of Servo and Spindle drive system.



Katsuhiko Kaji received a Ph. D. in information science from Nagoya University in 2007.
He became a RA at the NTT Communication Science Laboratories in 2007 and an assistant professor at Nagoya University in 2010. He has been an associate professor of the Faculty of Information Science, Aichi Institute of Technology since 2015. His research interests include indoor positioning and remote interaction. He is a member of IPSJ.

Katsuhiro Naito received a graduate degree in science and technology from Keio University in 1999, and M.E. and Ph.D. degrees in engineering from Nagoya University in 2001 and 2004, respectively, was an assistant professor at Mie University from 2004 to 2014 and was a visiting scholar in the Computer Science Department of the University of California, Los Angeles (UCLA) in 2011. Since 2014, he has been on the Faculty of the Information Science department of Aichi Institute of Technology. The research interests include wireless communication technology, sensor network systems, vehicular communication systems, Internet of Things (IoT) systems, overlay networks, and network protocols and architectures. He received IPSJ senior membership and IEICE senior membership in 2015 and received the IPSJ Nagao Special Researcher Award in 2016.

Naoya Chujo received his B.E. degree in applied physics and his M.S. degree in information science and his Ph.D. degree in electrical engineering from Nagoya University in 1980, 1982 and 2004. He joined Toyota Central R&D Labs. in 1982. He has been a professor at Aichi Institute of Technology since 2010. His research interests are in the area of embedded system and automotive electronics. He is a member of IEEE, IPSJ, IEICE, IEEJ, and Informatics Society.

Tadanori Mizuno received the B.E. degree in Industrial Engineering from the Nagoya Institute of Technology in 1968 and received the Ph.D. degree in Engineering from Kyushu University, Japan, in 1987. In 1968, he joined Mitsubishi Electric Corp. From 1993 to 2011, he had been a Professor at Shizuoka University, Japan. From 2011 to 2016, he had been a Professor at the Aichi Institute of Technology, Japan. Since 2016, he is an Affiliate Professor at the Aichi Institute of Technology, Japan. His research interests include mobile computing, distributed computing, computer networks, broadcast communication and computing, and protocol engineering. He is a member of Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, the IEEE Computer Society and Consumer Electronics, and Informatics Society.

Submission Guidance

About IJIS

International Journal of Informatics Society (ISSN 1883-4566) is published in one volume of three issues a year. One should be a member of Informatics Society for the submission of the article at least. A submission article is reviewed at least two reviewer. The online version of the journal is available at the following site: http://www.infsoc.org.

Aims and Scope of Informatics Society

The evolution of informatics heralds a new information society. It provides more convenience to our life. Informatics and technologies have been integrated by various fields. For example, mathematics, linguistics, logics, engineering, and new fields will join it. Especially, we are continuing to maintain an awareness of informatics and communication convergence. Informatics Society is the organization that tries to develop informatics and technologies with this convergence. International Journal of Informatics Society (IJIS) is the journal of Informatics Society.

Areas of interest include, but are not limited to:

Internet of Things (IoT)
Smart Cities, Communities, and Spaces
Big Data, Artificial Intelligence, and Data Science
Network Systems and Protocols
Computer Supported Cooperative Work and Groupware
Security and Privacy in Information Systems
Intelligent Transportation System
Distributed Computing
Multi-media communication
Information systems
Mobile computing
Ubiquitous computing

Instruction to Authors

For detailed instructions please refer to the Authors Corner on our Web site, http://www.infsoc.org/.

Submission of manuscripts: There is no limitation of page count as full papers, each of which will be subject to a full review process. An electronic, PDF-based submission of papers is mandatory. Download and use the LaTeX2e or Microsoft Word sample IJIS formats.

http://www.infsoc.org/IJIS-Format.pdf

LaTeX2e

LaTeX2e files (ZIP) http://www.infsoc.org/template_IJIS.zip

Microsoft Word™

Sample document http://www.infsoc.org/sample_IJIS.doc

Please send the PDF file of your paper to secretariat@infsoc.org with the following information:

Title, Author: Name (Affiliation), Name (Affiliation), Corresponding Author. Address, Tel, Fax, E-mail:

Copyright

Publisher

Address: Informatics Laboratory, 3-41 Tsujimachi, Kitaku, Nagoya 462-0032, Japan

E-mail: secretariat@infsoc.org

CONTENTS