1. Dariusz SAWICKI[1], 2. Jacek RYDZEWSKI[2], 3. Tomasz SZWELNIK[3], 4. Jakub KALISKI[3], 5. Józef MAMZA[3]

Warsaw University of Technology (1), CODI Technical Group (2), VOICELAB.AI (3)
ORCID: 1. 0000-0003-3990-0121; 2. 0000-0001-6103-6419; 3. 0000-0002-8581-4278; 4. 0000-0002-9893-9387; 5. 0000-0002-8789-3854

doi:.10.15199/48.2025.01.16

Modern intercom system in the COVID-19 period

Abstract. *The COVID-19 pandemic has forced many restrictions in public space. It is also seen in appropriate Human – Computer Interaction (HCI) solutions in the consumer devices space. We present the design of an advanced intercom system in which we tried to take into account modern HCI capabilities and also social aspects. A specific distributed control and measurement system was created where "tailor-made" solutions allowed for the correct implementation of the project. The usability assessment of the operating intercom system confirmed the social expectations of users.*

Streszczenie. *Pandemia Covid-19 wymusiła wiele ograniczeń w przestrzeni publicznej. Jest to również widoczne w realizacji interakcji człowiek-komputer (HCI) urządzeń konsumenckich. Przedstawiamy projekt zaawansowanego systemu domofonowego, w którym staraliśmy się uwzględnić współczesne możliwości HCI, a także aspekty społeczne. Powstał specyficzny rozproszony system kontrolno-pomiarowy, w którym rozwiązania „szyte na miarę" pozwoliły na prawidłową realizację projektu. Ocena użyteczności działającego systemu potwierdziła społeczne oczekiwania użytkowników. (**Nowoczesny system domofonowy w okresie pandemii COVID-19**).*

Słowa kluczowe: domofon, interakcja człowiek–komputer, sterowanie głosem, pandemia COVID-19.
Keywords: intercom system, human-computer interaction, voice control, COVID-19 period.

Introduction

Human-computer interaction (HCI) is one of the fields of computer science that has been developing spectacularly in recent years. On the one hand, this is due to the rapid development of electronic components technology (especially sensors) and the increase in performance of computers and controllers responsible for operating devices. On the other hand, the development of HCI is also the result of growing expectations of users. Since we noticed that the interface does not have to be complicated, we want to operate the devices in an intuitive, easy and pleasant way.

Live in COVID-19 period is very difficult. For the individual, for the local community, for the state. Many of us believed that the pandemic problems would remain on the pages of school readings. However, it is one thing to read Marquez's book "Love in the Time of Cholera", or "The Plague" by Albert Camus, and another – very different thing to live in a restricted country, experience tragedy, and relive the death of loved ones. Restrictions used today in different countries are very similar: a ban on leaving home, restrictions on movement, keeping a safe distance in interpersonal contacts, avoiding touching various objects in public space. We can wonder if, using modern HCI solutions, we can improve security, ease the consequences of restrictions, or at least try to make life easier for people in difficult times. Intercom systems are one of such devices of common use that create such an opportunity.

The intercom is a device used for many years in private homes, blocks of flats and housing estates, as well as in various industrial situations. It is a natural development of the doorbell and allows opening a door or gate, remotely. Adding a camera (video intercom) has significantly improved the safety of residents, allowing seeing the face of a potential guest.

It would seem, that further development of this device is not necessary – it is difficult to imagine the artificial development of such a simple solution. However, it turns out that social expectations are much higher and are constantly rising. New, advanced features are needed. And the desire to meet the needs arising from various forms of contemporary interpersonal contacts gives designers almost unlimited opportunities of development. On the other hand, known examples of various modern technologies used in HCI allow developing intercom systems. They lead them to truly advanced solutions and go out towards the residents' expectations.

Attempts to analyze social needs show how extensive functions in intercom systems can be useful. However, technical or price restrictions often stand in the way. The intercom is one of those devices that should be above all simple, cheap and reliable. And in addition, it should be easy (intuitive) to use. Achieving these effects with modern techniques requires some compromises or very advanced solutions.

The COVID-19 pandemic has forced many restrictions in public space. The need to maintain an appropriate level of security is also associated with appropriate HCI solutions. It turns out that it may be consistent with the solutions used or required in intercom systems.

The main purpose of the work is to present selected solutions of the video intercom structure, in the context of social expectations and aspects related to COVID-19. On the other hand, intercom system as a complex, distributed control and measurement system has been introduced. We presented a description of design problems from idea to successful implementation. Many market intercom systems are known, but the authors are not aware of any publications containing detailed analyzes of the solutions – which may have very practical significance. The authors are also unaware of published studies and analyzes of the usability of such systems, especially those considering social aspects during the COVID-19 pandemic.

HCI – a Spectacular Evolution of Possibilities

It can be assumed that in a few (a dozen or so) years, consumer devices will recognize (or simply will know) the expectations of recipients. They will decide how they work – by fulfilling our fancies without the need for special control. However, before this happens, we have the opportunity to control using many different forms of cooperation.

Most consumer devices are equipped with appropriate controllers to facilitate use. Clear and colorful screens allow controlling by touch, with various forms such as multi-touch and dynamic commands – touch with shift, are used [1]. In cases where it is possible, contactless control is used, e.g. using finger or hand gestures [2,3].

Control using eye gesture, and gaze tracking (oculography) is also very popular [4,5,6]. In this case, eye blinking / winking recognition is applied [7,8,9]. In recent years, there are also introduced control options with head gestures and body language [10,11,12]. Developed technologies are of particular importance for people with

disabilities, allowing them to use computers and other devices effectively [13,14,15]. The last method developed in recent years is voice recognition and voice control [16,17]. A method known for many years but requiring quite advanced computing equipment.

Restriction in COVID-19 Period vs Social Needs in Door and in-House Communication

In the COVID-19 period many studies and analyzes have been conducted showing the impact of a pandemic on our lives, many studies have been published. An example of a report directly related to HCI is "The Smart Audio Report 2020" [18]. This report is one more study confirming the extremely strong impact of COVID-19 on our lives. 77% of U.S. adults have had a change in their typical routine due to the outbreak of COVID-19. 54% of the U.S. population stated that they plan to go only to places that are necessary, or they feel there safe. 41% of population are staying at home, they are saying that they are not leaving unless in emergency. In new research the popularity of smart speakers and voice-assistants were analyzed. The result shows that such solutions increasingly are becoming a part of everyday lives for a large group of people. Usage of voice tech several times a day (or nearly every day) declares 52% users of voice-assistant now. Before the outbreak it was approximately 46%. 36% of U.S. adult owners of smart speaker say they are using their device more to listen to music and entertainment since the outbreak, and 52% of 18-34-year-olds say the same. 35% of U.S. adult owners of smart speaker are listening to more news and information since the COVID-19 outbreak, and 50% of those ages 18-34 say the same.

In COVID-19 period the natural need is the possibility of controlling devices in contactless way. Of course, we can use gloves, but it is better to adhere to the principle of "you will not touch, you will not get infected". Pandemic imposes some solutions. However, it seems that in this case it is surprisingly in line with HCI development trends. Today we can indicate many situations in which communication and voice control can be much more convenient and at the same time modern HCI solutions can provide it:

- The user belongs to group of disabled persons.
- The user's preference is to use natural language in communication, without first reading the application's instruction manual.
- The user has difficult access to the device, keyboard or screen.
- The user simply has his hands or eyesight busy.

The analysis carried out among the potential recipients of video intercom systems gave quite unequivocal indications regarding the clients' expectations towards such systems. Buyers on new markets (new construction investments) as well as old ones expect a coherent and uniform communication system and access control. One that not only works effectively but gives a sense of security and is friendly to use.

Intercom as a Modern HCI Tool

A door phone, in its basic form, is a simple intercom that allows communications between street (staircase, gate etc.) and house, giving the possibility to unlock / open doors, gates, etc. Modern solutions go far beyond this minimum. Especially, if we try to address the problem described in previous chapter. Taking into account these considerations and known sociological analyzes, we can try to formulate a set of social expectations, relevant for the recipients – users of intercom systems. These expectations can be divided into several groups:

- Increased the sense of security. This is one of the most important, if not the most important, social expectations. However, it is worth paying attention to the fact that the sense of security is identified with two aspects: passive and active. Passive, when expectation is access control - the possibility to analyze (also visually) who wants to get to an apartment, building, or a protected area. This should also be possible if the owner is outside his apartment, but would like to check, for example by phone, who is just calling the door. Active, when event response is required. In this case, it is important to be able to quickly call emergency services or the police through the intercom system.
- Widely understood functionality. Today, it is no longer enough to control the opening of a door or gate. All users using the Internet are accustomed to providing any information by any methods. Therefore, it is natural to expect the intercom system to allow contact as well with the child in the playground as with a neighbor in the building. So a wide range of "addressing" for contacts is very important. In addition, we expect the possibility of connecting to the mobile network in any situations. A classic example is the possibility of diverting an intercom call to a mobile phone in a situation when the courier delivers the package and the owner of the apartment is away from home.
- Friendly operation for elderly or disabled people. The natural solution expected by these groups of people is verbal communication and voice control of the devices.
- Easy installation. Preferably wireless.
- Reliability. This concept is understood today quite uniquely. Users usually expect maintenance to be easy, not disturbing and unnoticeable. This means expecting that the system administrator / maintenance person will do his work remotely, in a manner invisibly to the user. The possibility of remote checking, remote control and testing is therefore needed. And also, remote repairing, e.g. by replacing the device software.
- Contactless access. Possibility of controlling devices in contactless way. The voice recognition, verbal communication and voice control would be preferred.

Voice Control as a Contactless Solution

The most effective human-machine communication takes place today thanks to multimodal interfaces. By using many media, they allow using communication by sound, text and image, and gesture also. It is very important for man, because such communication is known for man and has been developed throughout his life. Instead of using buttons (or other more or less complex manipulators), people can use dialog systems where speech recognition is applied. This is especially important for disabled people and elderly people who notice many barriers of communication in a modern information environment.

In intercom systems, an important aspect is voice communication from a distance using microphones built into the device. The user does not need to interact directly with the device; thanks to speech recognition can enter an identifier, a password or any voice command from a distance. Examples of such applications are ALEXA assistants, Google Home, or other smart applications. According to "The Smart Audio Report 2020" [18], 24% of Americans over 18 use smart-speaker. The study was conducted in an American online survey of adults 18 years of age and older. It was carried out between March 31, 2010 and April 1, 2020, after the WHO declared the COVID-19 pandemic. It turns out, the number of people declaring the use of voice assistant several times a day in the period March 31 to April 1, 2020 increased

to 25% compared to the period from December 31, 2019 to January 5, 2020 when there were 20%.

Project of an advanced video intercom system

The aim of the project was to try to build an intercom system that would meet all social expectations discussed in previous chapters. It would seem, that modern electronic solutions in the field of control and communication do not create any barrier. At the same time, computer development and developed algorithms will facilitate this task. For a typical smartphone user, this device has today unlimited possibilities and can do almost everything. However, when we look more closely at selected applications, we often see solutions that perform advertised functions but in a very limited or selected range. This is most often the result of trade-offs between functional expectations and technical possibilities (and limitations).

It is worth analyzing the market of intercom systems manufacturers. There are many solutions at the highest world level. However, the question of whether they meet all expectations remains open.

To meet the project requirements, we decided to design it as a modular distributed system and divide the implementation into two independent tasks. The first is to solve technical aspects when handling communications, both internal in the system and with external objects. The second is voice support, which includes speech recognition and analysis of voice commands, followed by the formulation of system instructions.

Video Intercom – Technical Aspects

The progress of miniaturization has made the modern smartphone support many different communication systems. Similarly, to the expectations of intercom solutions. The comparison of a smartphone and intercom system modules is quite important because both devices can have similar sizes, and thus it is possible to similarly packing of electronic components. And similar computational capabilities of the processors used. However, what distinguishes devices of intercom system is a much higher, level of security and reliability.

From the point of view of communication and system service, several independent groups of tasks have been distinguished:

- Support of external devices. It is possible to connect almost any communication device (audio, camera, mobile phone, video monitor, etc.) to the line of resident. The possibility of any data transmission between these devices. It is possible to combine these devices into functional groups (for the purpose of distinguishing buildings or their parts).
- Intra-system communication. Full interconnection between all system modules (also between e.g. all residents within a large housing estate – without quantitative restrictions). Possibility of any data and address / identifier redirection as well as selection of any target connection.
- A set of basic functions of intercom system also including system administration and configuration, especially with personalization facilities.
- Support for additional security systems. Any communication with the service, administration, security, concierge etc. Handling emergency and alarm situations. Generation and transmission of notifications, handling and distribution of alarm and emergency information.

The System of Voice Control

Voice control is an integral solution of the designed system. It is worth emphasizing that the intercom system properties do not require universality in speech recognition. The goal is to recognize specific – selected commands and numbers in a proposed convention. Analyzing the system functions, a set of commands required to operate the devices was prepared. Additionally, the use of appropriate sensors should allow detection of a user approaching the intercom. Voice interaction using simple voice commands should allow you to initiate a simple dialogue. Ultimately, this will allow entering the apartment number and start access to the building. This way of interaction can encourage the elderly and/or disabled people, to use modern technologies. At the same time, it is in line with expectations in the COVID-19 period. The key problem here is not the universality of the analysis (and conducting a sophisticated conversation), but the effectiveness of speech recognition, which allows convenient use in various conditions, e.g. noisy acoustic background.

Implementation – the New Product

The entire introduced video intercom system has been developed in a modular form. Most of the system software was written in assembler as an individual, custom-made solution, optimizing the size and efficiency of the code. The system consists of the following modules:

- **The external panel** is supported by two MC9S08SH16 and SH8 processors. Includes a new generation QT60168 touch-sensitive keyboard controller with sensitivity of a femtofarad level. The software used (PE-Micro assembler) enables cyclical auto-calibration of the controller. Thanks to this high stability of keyboard is ensured. The panel also contains an RFID reader system based on the NXP self-tuning element – HTRC110 and a 128x64 graphic display designed and manufactured specifically for this solution.
- **The main module** controlled by the MC9S08PA64 processor containing audio path, video path (amplifier with auto-correction), higher voltage converter (external telephone standard). The software of this module contains of approx. 30 thousand program lines in assembler. The module has built a power supply for system of reversible lock with a custom controlled pulse transducer. The patented solution allowed reducing the power consumption four times in comparison to typical applications.
- **Commutator** is a commutating element of the multimedia signal (sound and image) in the bus between premises. It is a unique solution that allows transferring any signal between any destinations (modules). The solution has been patented. The commutator was built as a distributed shift register based on HEF4015 circuits and semiconductor relays.
- **The access control interface** is based on the MC9S08SH8 processor and 256kB EEPROM memory. It enables communication with the outside world via the Wiegand 26 protocol and via a serial interface with a system bus. It supports approx. 4000 RFID cards and signals from any sensors used in the system.
- **Video monitors** which are based on MC9S08SH16 processors and MC34118 half-duplex communication system. The monitors contain a special correction amplifier circuit (based on the Texas Instruments THS4221 element) that allows repairing a video signal, distorted during poor quality cable transmission.
- **Stand-alone voice communication module** based on Raspberry Pi3 computer with speech recognition and voice control software.

Voice Recognition and Voice Control

Using their own experience, the authors developed a speech recognition and voice control system for the intercom system as an especially dedicated one. It is worth noting that the adaptation in this case to specific applications is closely associated with the dictionary specificity and matching of language models. Banking, healthcare or insurance solutions will be different [19]. Individual, dedicated solutions of the authors were used, for example, in the banking industry to operate a bank account using a mobile application and a telephone line. This required the preparation of a dedicated dictionary in the banking industry and a language model with an acoustic model. Similarly, dedicated solutions were developed for the logistics industry.

In addition, the individualization of the solution allows transferring the software to virtually any computer platform. The authors' experience indicates the possibility of effective use of the same speech recognition algorithms and voice control on servers, desktops, mobiles and embedded systems.

We have developed our own speech recognition system based on deep neural network technology. Due to the modularity (embedded system) of the adopted solution, speech recognition and voice control software has been implemented in the Advanced RISC Machine (ARM) architecture on the Raspberry Pi platform.

In the first stage of speech recognition, two components can be distinguished: keyword detection and command recognition. Both of these tasks are implemented using the same acoustic model to reduce system response delay. The acoustic model uses recursive neural networks Long Short-Term Memory. The network was trained with application of the Connectionist Temporal Classification cost function. In order to provide the desired speed of operation on the embedded system, neural network calculations are subject to significant quantization and use Single Instruction Multiple Data operations specific to the ARM architecture.

The second stage is matching the phrase based on the probabilities of the acoustic model. most commonly, the language model is used for this, which is a very general solution that requires significant resources. In the case of voice control in the intercom system, the statement has a known structure. This allowed us to propose a better solution, which is the Speech Recognition Grammar Specification technique. We used stored, structured statements that allow you to determine what to expect in the recognized audio signal. In word detection, we used a greedy algorithm to evaluate phrases based on the probabilities of an acoustic model. In addition, the GARBAGE rule was applied to improve recognition quality. This rule returns information about a statement that does not match the declared grammar. This avoids false detections/recognitions.

The Final Effect: Tests of the System

The reliability of the system operation is one of those features of the intercom system, which is very desirable. To ensure a sufficiently high level of reliability, a set of tests was carried out under three categories: laboratory tests, tests in real conditions, usability evaluation. The first two categories of tests allow us to check the correctness of the system construction. Correctness of the design and implementation. Both in specialized/test conditions and during standard work in real conditions. Third category tests allow us to assess the quality of the interaction proposed for users to operate the system. Since the intercom system should be a simple, useful device that does not cause any problems to users, this assessment is very important, especially in times of pandemic.

Laboratory Tests

These tests included independent software and hardware implementation. Testing took place in two different situations. During "normal" operation, where the system behavior was checked in typical situations. And during "atypical" work – such tests included a prepared set of emergency and unusual situations – resulting from incorrect operation, incorrect installation and / or incorrect assembly. These types of tests particularly concerned software solutions to avoid system freezes due to incorrect operation by users.

After conducting several tests, we also introduce to apply self-analysis of correctness and self-correction of the system. The entire system software includes a set of test procedures that are carried out on a regular basis from time to time (defined frequency). Thanks to this, an analysis of system correctness is conducted. If errors are detected, the corresponding module is corrected or reset. Event log allows monitoring the work over a longer period of time (as a maintenance task for the administrator).

Tests for resistance to electromagnetic interference from outside were also carried out. Climatic tests were carried out – e.g. the external panel was tested at temperatures from -25ºC to +60ºC. Added to this are moisture tests and rain resistance.

Selected modules (in particular an external panel with a touch keyboard) were tested for the impact of power disturbances on the possibility of spontaneous generation of access control signals (door opening / closing). After many attempts, an appropriate auto reset solution has been proposed that completely eliminates the possibility of this type of work disorder. This increases the security of system operation. This solution has also been patented.

Tests in Real Conditions

The tests were carried out at the selected facility. Over a period of several weeks, the system was evaluated by residents. After completing the tests and analyzing the collected information, software corrections were made to facilitate the operation of the devices. Most often, these corrections related to certain time dependencies that are preferred by the residents.

We have tested the speech recognition and voice control module independently. A dedicated acoustic model was developed and sets of dictionaries prepared. Thanks to this, the system ensured high recognition efficiency independent of the speakers. Samples used for training were prepared on the basis of acoustic signals collected from statements of many thousands of people.

The Word Error Rate (WER) measure is used to determine the quality of the speech recognition system [20]. It is a measure that determines the recognition error at the word level, which compares the recognition pattern with the recognition received from the speech recognition system.

$$WER = \frac{S+D+I}{N} = \frac{S+D+I}{S+D+C} \tag{1}$$

where:
S – the number of substitutions, D – the number of deletions, I – the number of insertions, C – the number of correct words, N – the number of words in the reference ($N=S+D+C$).

For the English test set LibriSpeech, subset dev-clean [21,22], on the general language model for the continuous speech recognition system, the WER is about 7.43%. The use of proper grammars improves this result, while allowing speech recognition on embedded devices with limited resources.

A set of tests was carried out on a separate group of samples. For the dedicated solution, the WER result was

4-5% depending on the degree of background noise. Such a result allows finding that the system works correctly at a level of at least 95%.

Usability Evaluation

We conducted a usability evaluation of the intercom among residents using this system. Two surveys were prepared, which were completed by users. The evaluation was voluntary. The surveys were completed and submitted by residents who wanted to voluntarily share their opinion. 20 participants took part in the evaluation.

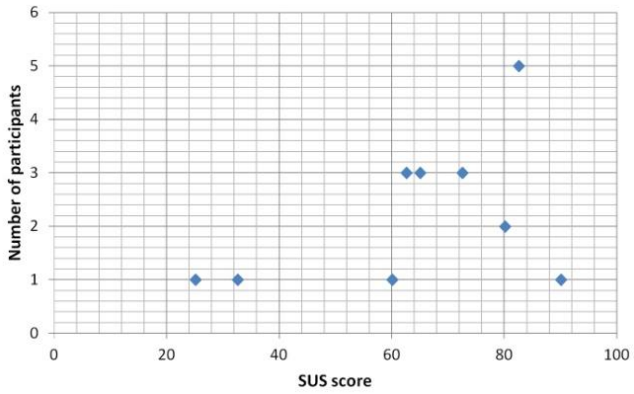

Fig. 1. SUS score in the first evaluation.

The first assessment was performed using the System Usability Scale (SUS), introduced by Brooke in 1996 [23]. The SUS questionnaire contains 10 questions. The response to each is a choice of acceptance level, according to the five-point Likert scale [24], from "strongly disagree" to "strongly agree". The Polish version of the questions (APPENDIX) has been developed. Taking into account the calculation rules of the results for SUS, each participant could finally assess usability on a scale from 0 (worst) to 100 (best) points. The ratings of individual participants (Table 1) are presented in Figure 1. We interpreted the results in accordance with [25] in two categories: Adjective Rating (Table 2) and Acceptability Rating (Table 3).

2 participants (out of 20) described the system as Awful and the same 2 people declared it Not Acceptable. This represents 10%. At the same time, more than half (11 out of 20) described the system as at least Good and also 11 out of 20 unambiguously accepted the system. This is 55% of participants. 35% of participants partially accepted, having reservations about selected problems. Analyzing the results, it can be concluded that the assessment of the system by users using SUS is not unambiguous. However, the vast majority assessed the system positively.

The second assessment of the system was based on the ISO 9241–411:2012 standard [26]. Questionnaire consisted of 5 questions. The first three were related to technical issues and were primarily related to the use of voice control. The last two questions concerned social problems. The response to each is a choice of acceptance level, according to the five-point Likert scale [24], from "strongly disagree" to "strongly agree". The Polish version of the questions and the English translation (APPENDIX) has been prepared. The results of participants' assessment are presented in Figure 2.

Table 1. Result of the first test: (for 20 participants)

Participant	Evaluation (SUS score)	Participant	Evaluation (SUS score)
1	72.5	11	25
2	80	12	65
3	62.5	13	32.5
4	82.5	14	62.5
5	77.5	15	87.5
6	82.5	16	72.5
7	65	17	65
8	87.5	18	90
9	62.5	19	60
10	87.5	20	80

Table 2. Adjective Rating scale in SUS test of our system

Range of points	Adjective Rating	In our test
12.5 – 20.2	Worst Imaginable	0
20.3 – 35.6	Awful	2
35.7 – 50.8	Poor	0
50.9 – 71.3	Ok	7
71.4 – 85.4	Good	10
85.5 – 90.8	Excellent	1
90.9 – 100	Best Imaginable	0

Table 3. Acceptability Rating scale in SUS test of our system (for 20 participants)

Range of points	Acceptability Rating	In our test
below 50	Not Acceptable	2
50 – 62.6	Low Marginal	4
62.6 – 70	High Marginal	3
above 70	Acceptable	11

Analyzing the results of the second assessment, it can be stated that the assessment of the system is differentiated in an interesting way. Users accept and evaluate the proposed technical solution well – the first three questions (average above 4). On the other hand, they negatively assess the impact of the system on social relations (both in terms of neighborly and family relations) – the last two questions (averages 3.15 and 2.15). However, it seems that in this case the general (and bad) assessment of the pandemic and COVID-19 period prevailed. And even the best technical solutions could not change it.

Conclusions

The paper presents the project of a specific distributed control and measurement system in HCI application. The developed intercom system has been designed as a modular

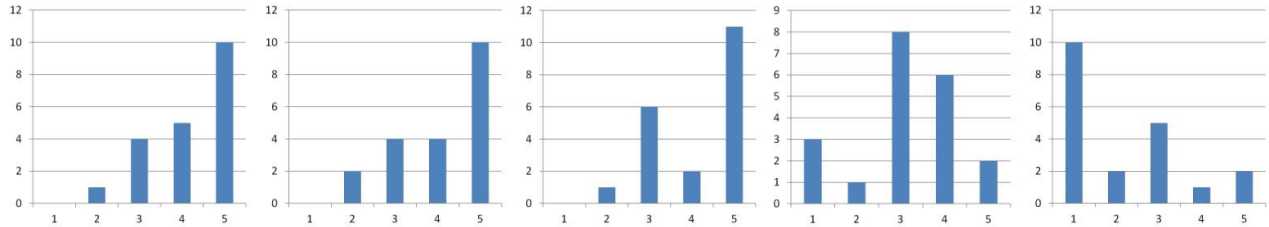

Fig. 2. Frequency of grades in the participants' evaluation from 1 ("strongly disagree") to 5 ("strongly agree") in the questionnaire based on the ISO 9241–411:2012 standard. The graphs show the number of particular answers for five questions (graphs for question from left to right). The first question, the average result was 4.2 (δ = 0.96). The second question, the average result was 4.1 (δ = 1.08). The third question, the average result was 4.15 (δ = 1.05). The fourth question, the average result was 3.15 (δ = 1.19). The fifth question, the average result was 2.15 (δ = 1.42).

system while optimizing the solutions of each of the components and at each stage of construction. Each module uses specific dedicated solutions. This approach required a lot of work on the individualization of the project but allowed to achieve the expected effect of the whole. Introduced project is an example of situation, where only „tailor-made" solutions allow achieving final success.

This solution uses dedicated speech recognition and voice control systems. Recognition methods are based on deep neural network technology. This is a known approach that produces good results. However, the advantage of the method used in this case is the analysis resulting from the needs and expectations of the project. Thanks to this, the operation of the neural network was adjusted, which significantly improved the efficiency of the system. As a result of individual parameterization for the needs of a particular application, a solution was obtained that exceeded the solutions available on the market.

The project was carried out in 2018-2019. In 2020 implementation took place. This last stage has coincided with the COVID-19 pandemic. The usability tests were done in 2022 – the system has been well evaluated by users. The analysis of project requirements turned out to be convergent with the requirements for operating devices in COVID-19 period. For some people this may come as a surprise. However, it seems that looking at the development of HCI; it is worth noting some obvious aspects. The development of HCI is a compromise between usability and security, and technical capabilities. Even *science-fiction* solutions don't lead only to more interesting effects in computer games. Sooner or later, this will lead to increased usability and security in common device operation. Increasing safety is a decisive element of broadly understood usability in HCI. As in the case of intercom systems and in the described project.

APPENDIX

The first assessment.

The original SUS questionnaire [23]:

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The Polish translation of SUS questionnaire:

1. Myślę, że chciałbym często korzystać z tego systemu.
2. Według mnie system jest niepotrzebnie skomplikowany.
3. Odniosłem wrażenie, że system jest prosty w obsłudze.
4. Myślę, że potrzebuję wsparcia osoby z doświadczeniem technicznym, aby móc korzystać z tego systemu.
5. Zauważyłem, że różne funkcje tego systemu są dobrze zintegrowane.
6. Miałem wrażenie, że w tym systemie jest zbyt wiele niekonsekwencji.
7. Wyobrażam sobie, że większość ludzi bardzo szybko nauczyłaby się korzystać z tego systemu.
8. Uważam, że system jest bardzo niewygodny w użyciu.
9. Czułem się bardzo pewnie korzystając z tego systemu.
10. Musiałem nauczyć się wielu rzeczy, zanim mogłem zacząć korzystać z tego systemu.

The second assessment

The original Polish questionnaire.

1. Dzięki sterowaniu głosem wybieranie odpowiednich funkcji jest szybkie i precyzyjne.
2. Sterowanie głosem poprawia komfort obsługi domofonu.
3. Zamiana sterowania dotykowego na głosowe poprawiła bezpieczeństwo pracy w czasie pandemii.
4. Realizacja interfejsu domofonu przyczyniła się do poprawy komunikacji społecznej i relacji sąsiedzkich.
5. Korzystanie z domofonów wpłynęło korzystnie na relacje rodzinne w czasie pandemii.

The English translation

1. With voice control, selecting the right functions is quick and precise.
2. Voice control improves the comfort of using the intercom.
3. Replacing touch control with voice control improved work safety during the pandemic.
4. The implementation of the intercom interface contributed to the improvement of social communication and neighborly relations.
5. The use of intercoms had a positive impact on family relationships during the pandemic.

***Authors**: Prof. Dariusz J. Sawicki PhD. DSc., Warsaw University of Technology, Institute of Theory of Electrical Engineering, Measurement and Information Systems, Koszykowa 75, 00-662 Warsaw Poland, E-mail: dariusz.sawicki@pw.edu.pl. Jacek Rydzewski MSc. Eng., CODI Technical Group, Renesansowa 7C, 01-905 Warsaw Poland, E-mail: elita@codi.pl. Tomasz Szwelnik MSc. Eng., Jakub Kaliski MSc. Eng., Józef Mamza MSc. Eng., VoiceLab.AI, Grunwaldzka 135A, 80-264 Gdańsk, Poland. E-mails: Tomasz@voicelab.pl, jakub.kaliski@voicelab.pl, jozef.mamza@voicelab.pl.*

REFERENCES

[1] Oviatt, S., Cohen, P.R., 2015. *The Paradigm Shift to Multimodality in Contemporary Computer Interfaces*. Morgan & Claypool.

[2] Krupka, E., et al. (2017). Toward Realistic Hands Gesture Interface: Keeping it Simple for Developers and Machines. In: *Proc. 2017 CHI Conference on Human Factors in Computing Systems*. Denver USA. pp. 1887-1898. DOI: https://doi.org/10.1145/3025453.3025508

[3] Wachs, J. et al., (2006). A Real-Time Hand Gesture Interface for Medical Visualization Applications. In: Tiwari A., Roy R., Knowles J., Avineri E., Dahal K. (eds) *Applications of Soft Computing. Advances in Intelligent and Soft Computing*, vol 36. Springer, Berlin, Heidelberg. DOI: https://doi.org/10.1007/978-3-540-36266-1_15

[4] Duchowski, A., (2007). *Eye tracking methodology. Theory and practice*. sec. ed. Londyn: Springer.

[5] Singh, H., Singh, J. (2012). Human eye tracking and related issues: a review. *International Journal of Scientific and Research Publications*. 2(9), 1-9.

[6] Kwiatkowska, A., Sawicki, D., (2018). Eye Tracking as a Method of Controlling Applications on Mobile Devices. In: *Proc. of the 15th International Joint Conference on e-Business and Telecommunications (ICETE 2018) – Volume 1: DCNET, ICE-B, OPTICS, SIGMAP and WINSYS*, Porto, Portugal. pp. 373-380. DOI: http://dx.doi.org/10.5220/0006837003730380.

[7] Majaranta, P., Bulling, A., (2014). Eye Tracking and Eye-Based Human–Computer Interaction. In: Fairclough, S., Gilleade, K. (eds) *Advances in Physiological Computing. Human–Computer Interaction Series*. pp. 39-65. Springer. DOI: http://dx.doi.org/10.1007/978-1-4471-6392-3_3

[8] Singh, H., Singh, J., (2018). Real-time eye blink and wink detection for object selection in HCI systems. *Journal on Multimodal User Interfaces*, 12(1), 55-65. DOI: http://dx.doi.org/10.1007/s12193-018-0261-7

[9] Kowalczyk, P., Sawicki, D., (2019). Blink and wink detection as a control tool in multimodal interaction. *Multimedia Tools and Applications*. 78(10), 13749–13765. DOI: http://dx.doi.org/10.1007/s11042-018-6554-8.
[10] Evans, D.G., Drew, R., Blenkhorn, P., (2000). Controlling Mouse Pointer Position Using an Infrared Head-Operated Joystick. *IEEE Transactions on Rehabilitation Engineering*, 8(1), 107–117.
[11] Kim H., Ryu D., (2006). Computer control by tracking head movements for the disabled. In: *Proc. of the ICCHP '06*. In: *Lecture Notes in Computer Science*, 4061, pp.709–715, Springer.
[12] Sawicki, D., Kowalczyk, P., (2018). Head Movement Based Interaction in Mobility. *International Journal of Human–Computer Interaction*. 34(7), 653-665. DOI: http://dx.doi.org/10.1080/10447318.2017.1392078.
[13] Strumiłło, P., Pajor, T., (2012). A vision-based head movement tracking system for human-computer interfacing. In: *Proc. New trends in audio and video/signal processing algorithms, architectures, arrangements and applications (NTAV/SPA)*. Łódź, Poland. pp. 143-147.
[14] Szczepaniak, O., Sawicki, D., (2017). Gesture controlled human–computer interface for the disabled. *Medycyna Pracy*. 68(1), 11-21. DOI: http://dx.doi.org/10.13075/mp.5893.00529
[15] Guzsvinecz, T., Szucs, V., Sik-Lanyi, C., (2019). Suitability of the Kinect Sensor and Leap Motion Controller-A Literature Review. Sensors (Basel). 19(5), 1072. DOI: http://dx.doi.org/10.3390/s19051072.
[16] Tandel, N.H., Prajapati, H.B., and Dabhi, V.K., (2020). Voice Recognition and Voice Comparison using Machine Learning Techniques: A Survey. In: *Proc. 6th International Conference on Advanced Computing and Communication Systems (ICACCS),* Coimbatore, India, 2020, pp. 459-465, DOI: 10.1109/ICACCS48705.2020.9074184.
[17] Singh, A.P., Nath, R., and Kumar, S., (2018). A Survey: Speech Recognition Approaches and Techniques. In: Proc. *5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Gorakhpur, 2018, pp. 1-4, DOI: 10.1109/UPCON.2018.8596954.
[18] The Smart Audio Report, (2020). National Public Media, USA, Report, April 2020. available at https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/ last accessed 14 May 2022.
[19] Intent Recognition Use Cases, Use VoiceLab to improve your business service quality in any industry. https://voicelab.ai/#use-cases/ last accessed 10 January 2022
[20] Ali, A., Renals, A., (2018). Word Error Rate Estimation for Speech Recognition: e-WER. In. *Proc. 56th Annual Meeting of the Association for Computational Linguistics*. July 2018, Melbourne, Australia. Volume 2: pp.20-24. DOI: http://dx.doi.org/10.18653/v1/P18-2004
[21] Panayotov, V., Chen, G., Povey, D., Khudanpur S., (2015). Librispeech: An ASR corpus based on public domain audio books. In: *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 16-24 April 2015, Brisbane, QLD, Australia, pp. 5206–5210. DOI: http://dx.doi.org/10.1109/ICASSP.2015.7178964
[22] Open SLR Resources, (2023). Available at: http://www.openslr.org/resources.php/ last accessed 20 October 2023
[23] Brooke, J., (1996). SUS: A 'Quick and Dirty' Usability Scale. In: *Usability Evaluation in Industry*, Jordan, P.W., Thomas, B., McClelland, I.L.,Weerdmeester, B., (Eds.), London, Taylor and Francis, pp. 189–194.
[24] Likert, R., 1932. A technique for the measurement of attitudes. *Arch. Psychol*. No 140, 5-55.
[25] Bangor, A.; Kortum, P.T.; Miller, J.T., (2008). An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact*. 2008, 24, 574–594.
[26] ISO Standard: ISO 9241-411:2012 (2012) Ergonomics of human-system interaction – Part 411: Evaluation methods for the design of physical input devices