Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 6169-6174

Research Article

Development of Kodály Hand Sign Educational Tools using Multi-Layer Perceptron

Young Bae Yun [a], Young Man Kwon [b], Myung Jae Lim [b], Jeong Jun Park [c] and Dong Kun Chung [d]

[a] Professor of Department of Child Studies, Eulji University, Seongnam.
[b] Co-Corresponding Author, Professor of Department of Medical Information Technology, Eulji University,
[c] Undergraduate of Medical Information Technology, Eulji University
[d] Corresponding Author, Professor of Department of Medical Information Technology, Eulji University, tchung@eulji.ac.kr

Abstract: Hand signs have been recognized to be useful in enhancing the effectiveness of music learning in music education. In particular, Kodály hand signs based on the Kodály teaching method are widely used around the world because their effectiveness has been recognized. The purpose of this study is to implement a hand sign recognition system using a neural network and to present a method that can systematically learn hand signs. Device-based systems were considered for portability, activity, and economy. In addition, the introduction of additional filtering to the neural network input was proposed to improve portability and recognition rate to small devices.

Keywords: Kodály hand signs, neural network, motion recognition, gesture recognition, MLP

1.Introduction

During the development of children's musicality, the sense that develops later than the sense of rhythm and has the greatest individual difference is the sense of tone that distinguishes the height of the tone **(YoungBae.Yun.2019)**. Among the methods for teaching children's sense of sound, there is a method of using hand signs. The hand sign system was created to help church choir members as well as student training in school for more accurate recognization of the pitch.

The Hand Sign originated from the fact that Guido D'Arezzo (995~1050), an Italian monk and music theorist, used the palm of his hand to enhance the educational effect of the choir. After that, English music educator John Spencer Curwen (1847-1916) taught the names of the hands using various shapes of hands, and today the prototype of Hand Sign was born. In teaching beginners, including young children, Curwen systematized singing education in school by avoiding the system of staff and using hand signs with the names written on them. Hand Sign used in music education for children and children today, was revised and supplemented by Hungarian composer and music educator Zoltan Kodály (1882-1967 Hungary) **(HongGi.Cho.1994) (WanByeong & HongGi 1999)**. The Kodály hand sign has been created and disseminated into a more specific and functional hand sign, with the addition of 'Pa', 'Ti' and semitone divisions, and is currently used in song education.

The purpose of this study is to develop a more effective educational tool by fusing these Kodály hand signs with Neural Network technology. Activity, portability, and economy were considered in the system design to increase its utility value as a useful educational tool for infant singing education and music activities in general. In connection with the recognition of hand signs, a multi-layer perceptron was introduced and a study was conducted to improve the recognition rate.

2.Related studies

2.1.Kodály Teaching Method

Zoltán Kodály (1882-1967) is a Hungarian composer and music educator. He insisted that 'music should belong to everyone, not just a few selected upper-class people. In particular, he said that all children who can hear and speak can improve musicality, so music education should be implemented systematically in stages.

Kodály considered that it is effective for infants to start learning intervals with minor 3rd and major 2nd intervals in terms of development. Most of the songs and chants are sung by infants are composed of three sounds: sol-mi and sol-mi-la. Singing instruction allows you to first experience the auditory experience by listening to the teacher's song and singing along with the hand sign syllable names. Next, notations are made with a staff made of felt cloth on the velvet and a staff made of magnet on the magnetic board. While reading the notated music sheet visually, the ability to read sheet music is enhanced through the three-step learning process of listening to songs, writing, and reading again **(Kodály.Zoltán.1982) (Kodály.Zoltán.1974)**.

The advantages of using hand sign are as follows: First, the pitch is clearly visible in space. second, the concept of each note is specifically formed for the pitch and name to be expressed. third, sound function can be shown in detail (Do-Tonic/ Wave-always going down/ Ti-always going up)/ Other shapes and properties of semitones). fourth, Inner Hearing training to sing and listen to the sound in your mind. fifth, it can bring out the sound through the inner hearing without the help of a piano or instrument. sixth, it can practice two parts for polyphony training, and it is good for intonation training. seventh, The cerebral function responsible for the body's moving function and the cerebellar function responsible for intelligence work together **(YoungBae.Yun.2019)**.

The hand sign of the Kodály teaching method is also used to indicate transposition. For example, if the melody of a song ends with a Sol in C major, one hand represents the hand sign of Sol, while the other hand represents the hand sign by replacing Sol in C major with a Do of another composition (eg, G Major). If so, Sol in C major becomes DO in G major. The hand sign is useful for teaching not only the pitch but also the name of the note. In fact, in terms of music education, it can be used not only for high-level students but also for enhancing the learning effect without being restricted to any age group **(YoungBae.Yun.2019)**.

In addition to the Kodály teaching method, the use of Hand signs in contemporary children's music education, such as the Dalcroze teaching method and the Orff Approach, has been widely used worldwide as its effectiveness has been recognized.

Dalcroze's solfège teaching method is an activity to sing with syllable names. This is an activity that learns to read syllable names using the Fixed Do System and adjusts with the interval. The main purpose is to learn the relationship between tonality and association and to develop hearing and inner hearing ability. By reading syllable names with the Fixed Do System, you can learn Absolute hearing and accurately grasp the tone pitch **(EunYoung.Lim.2013)**.

Singing of the Orff Approach is also systematically taught. At first, like the teacher and Zoltan Kodály, the children or children respond to each other or imitate the teacher's song with a sol-mi going down minor 3rd interval note. After learning to respond with a song using one's name and the object by making a simple melody, the sound 'Ra' is added, followed by 'Do' and 'Le'. In this case, the hand sign is usefully used **(HongGi.Cho.1994) (WanByeong & HongGi 1999)**.

After mastering the movable Do system of this Fantatonic sufficiently, use the Heptatonic scale that contains 'Fa' and 'Ti'. Singing is the basis of music class, and the voice is a physical instrument that everyone has without burden **(YoungBae.Yun.2019)**.

2.2.Hand sign system

The following Table.1 shows the hand symbols used in children's music education with some functional modifications from Curwen to Kodály. In Curwen's method, Kodály is Fa (pa) with thumb down, Temporary signature Fi (Pi-Pa♯) with thumb up, Ti (tee) by moving forefinger up and down., Ta (ta-shi♭) was indicated. In this paper, Do, Le, Mi, Pa, Sol, La, and Ti will be used.

3.Hand sign recognition system

Hand movements can be recognized as hand signs through various recognition methods. There are two ways to recognize hand movements. One is a vision-based method that utilizes infrared and video processing, and the other is a device-based method that utilizes sensor measurements **(HwangSoo & HyunTak 2020) (JunCheol.Lee.2018)**.

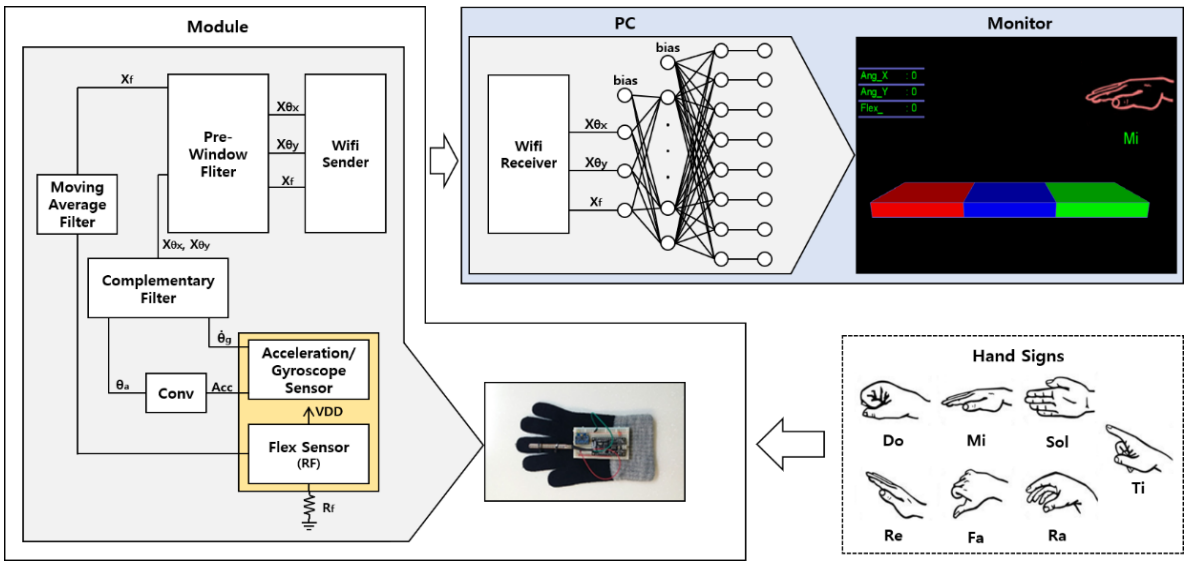
Vision-based methods that utilize infrared or video processing are convenient because they do not require the wearing of equipment for measurement. However, it has the disadvantage that it requires a separate fixed camera to check its operation and a high-performance computer to process the images. In addition, the user must also be in the field of the camera and there must be no obstacles between the hand and the photographic device **(JuYong.Jang. et al.,2014) (YoungMan.Kwon et al.,2019) (Gyöngyössy.N.M. et al.,2019)**.

In this study, a device-based system was considered for portability and activity while ensuring real-time performance. Figure.1 shows the system architecture.

Table.1. Kodály hand sign example

syllable names	Specification	function	Hand shape
Do'	-Hold the fist lightly and point it to the side at eye level. -It looks like the 'do' below, but raises	Above the octave key note	
Re	-Open the palm of your hand and face it up at an angle at the level of the belly button.	The first note from Tonic to the top	
Do	-Hold the fist lightly and point it to the side at the height of the belly button.	Tonic	

Figure.1. System Architecture



Compared to infrared and video processing, the method of using sensors to recognize hand movements not only requires much less processing power for computers but also reduces production costs. Furthermore, when using a wireless communication system such as WiFi or Bluetooth, there is an advantage that it can be expanded to all devices capable of wireless communication such as PCs and smartphones. To reflect the above, the design includes the Espressif ESP8266-12 module, which has an economical price as an MCU that supports Direct WiFi.

Each syllable name represented by hand signs is distinguished by the tilt due to the rotation of the X-axis, the tilt due to the rotation of the Z-axis, and the bending of each knuckle with respect to the right-hand coordinate system. A 6 DOF (Degree Of Freedom) module sensor (MPU60503-Axis Acceleration Gyroscope) was used to detect these tilts.

A Flex Sensor was used to measure the degree of knuckle bending. The value of resistance (Rf) of this Flex Sensor changes proportionally according to the degree of bending. This Flex Sensor, Rf, is connected in series with a fixed resistor (RF) and is connected between the power supply voltage and ground. Depending on the bending of the knuckle, a variable distribution voltage between the two resistors is applied to the input of the MCU's ADC and is used for hand signs recognition.

The bending degree value processed by the MCU is determined by Eq.(1).

The optimum value of the fixed resistance R_F in Eq.(1) is determined by ($R_F = \sqrt{R_{f_{max}} R_{f_{min}}}$), where ($R_{f_{max}}$) is the resistance value when the Flex Sensor is bent to the maximum, and Equation ($R_{f_{min}}$) is the resistance value when no external force is applied.

$$x_f = R_F K/(R_f + R_F) \text{ , where } K = 1024 \cdot V_{DD} \qquad (1)$$

Acceleration and angular velocity in the x, y, and z directions sensed from the 6DOF module are converted to the slope (angle) using an algorithm coded inside the MCU. In this process, a complementary filter algorithm was included in the program to obtain stable angles $x_{\theta x}$, $x_{\theta y}$, $x_{\theta z}$ with hand movements. (Eq.(2)).

In Eq.(2), $\theta_c = [x_{\theta x}, x_{\theta y}, x_{\theta z}]$, where $\dot{\theta_g}$ is the x, y, z angular velocity of the gyro sensor, and θ_a is the angle calculated from the acceleration value obtained from the sensor.

$$\theta_c = \frac{K}{s+\tau}\dot{\theta_g} + \frac{\tau}{s+\tau}\theta_a \qquad (2)$$

MLP(Multilayer perceptron), one of the artificial neural networks, was applied to the learning and recognition of hand signs. The input stage of the MLP is composed of the X and Y axis angles ($x_{\theta x}$, $x_{\theta z}$) and the degree of bending (x_f).

The hidden layer was designed with one stage, and the output stage consisted of eight nodes(Do, Re, Mi, Fa, Sol, La, Ti, 'silence'). In addition, the input stage and the hidden layer were configured to have one bias each.

In order to obtain the optimum conditions for recognition, 3 to 12 neurons were placed in the hidden layer, and then each test was performed. In addition, each weight (w) was trained by applying a backpropagation algorithm.

The weight is adjusted using Eq.(3) to which the Delta Rule is applied.

$$\Delta w_{ij} = \eta\delta_i x_j \ , \Delta w_{ij} = w_{ij} + \Delta w_{ij} \ , x = [x_{\theta x}, x_{\theta z}, x_f] \ , \eta : \textit{learning-rate} \ (3)$$

It was investigated how the characteristics of the input data formed by hand signs could affect the recognition rate, which led to the direction of improvement of the system. This was evaluated using K-fold cross validation.

4.Performance

Figure. 2 shows the hand-shaped data for each syllable name in a three-dimensional graph, where the input data of hand signs was measured 100 times every each syllable name. As shown in Figure. 2, when the measured data is classified within a specific range, the corresponding syllable name can be determined. However, in the real world, it is not easy to classify rapidly changing data between two-syllable names during musical activity. Therefore, it is necessary to exclude the data from the classification during the period of rapid change.

Figure. 3 shows the x1, x2, and xf data measured continuously every second by the movement of hand signs. (The order is "Do", "Re", "Mi", "Fa", "Sol", "La", and "Ti".)

In order to investigate the effect of rapidly changing data on the recognition rate, an evaluation was conducted by applying the data in Figure. 3 to the system in Fig. 1. K-fold cross-validation was used as the validation method, and fold was selected as 5. In Eq. (3), the learning rate was fixed at n = 3. Figure. 4 shows the case where epochs are set to 100 to 1000. From the figure, it can be seen that the increase in epoch does not lead to an increase in the recognition rate. Fig. 5 shows the case where the number of neurons on the hidden layer is set to 3 to 20 in the state of being fixed at epoch = 500.

Figure.2. syllable name & data by hand signs

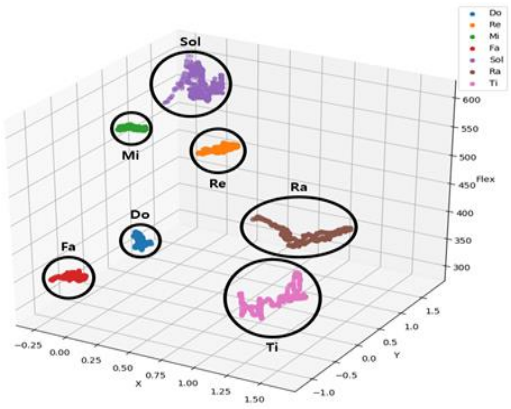
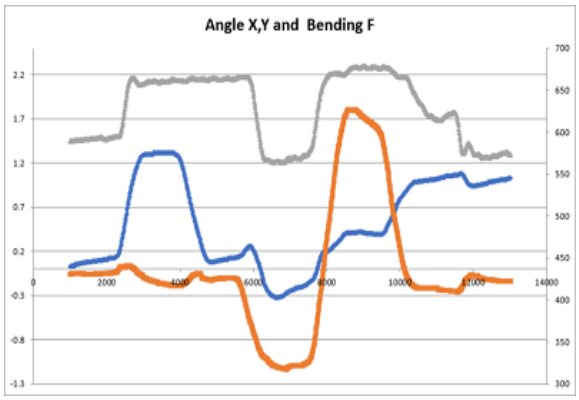


Figure.3. data by hand movements



When the number of neurons is 3, the recognition rate is 76%, and as the number of neurons is increased to 4,5, the recognition rate is also increasing at 83.6% and 86.8%. However, when the number of neurons is 9 or more, it only converges to 95%, and it is not possible to expect a higher recognition rate.

The reason why such a phenomenon occurs is considered to be due to the data measured between syllable names, as described above. These data should be classified as "silent", but in the real world, the recognition rate can only

be further reduced due to the data in all cases occurring between the seven-syllable names. In this case, if it is designed to go through a pre-filtering process for movement changes beyond a certain range, it will be able to not only improve recognition but also reduce the number of neurons in the hidden layer.

Figure.4. epochs and Recognition rate

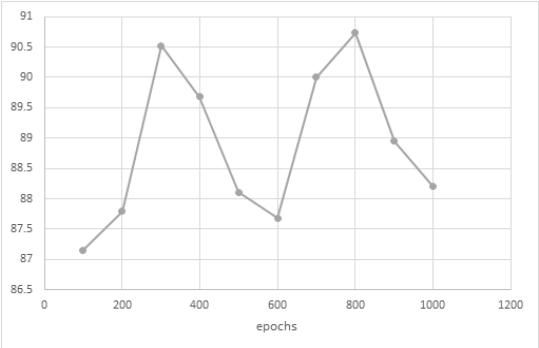
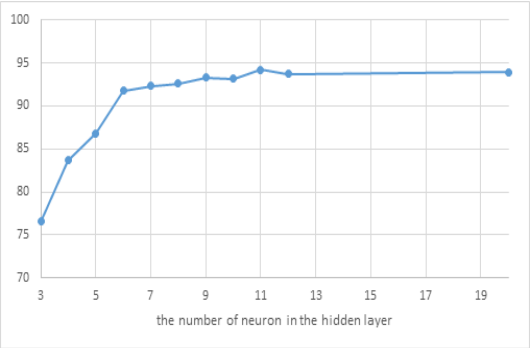


Figure.5. No. of neurons and Recognition rate



Eq. 4 can be considered for pre-filtering.

$$\frac{\Delta x}{\Delta t} = \frac{x_k - x_{k-1}}{\Delta t} \rangle \alpha \Rightarrow \alpha = [\, a_{\theta x}, a_{\theta z}, a_f \,]$$
$$if\ a_{\theta x} \rangle c_{\theta x}\ OR\ a_{\theta z} \rangle c_{\theta z}\ OR\ a_f \rangle c_f, then\ x_c = [\, k_{\theta x}\ ,\ k_{\theta z}\ ,\ k_f \,] \quad (4)$$

Table 2. shows the number of neurons and the recognition rate by epoch when the input value to the neural network circuit is set to xc = [0.7, 0.5, 625] each time a sudden change occurs. (Set learning-rate to 0.3 and the folds to 5).

Table.2. Recognition rate of proposed system

hidden neuron	epoch	time(sec)	Recognition(%)
3	500	60.4	97.5
3	700	83.4	97.7
3	1500	88.6	97.2
4	500	69.7	99.7
4	700	105.0	99.8
4	1000	139.6	100.0

As shown in Figure 1, by pre-filtering the input data that does not affect recognition, not only can the recognition rate be significantly improved, but also the number of neurons in the hidden layer can be reduced.

5.Conclusion

The purpose of this paper is to develop a performance device based on Hand signs recognition as a useful teaching material tool for infant singing education and music activities in general. Device-based development is being considered for portability and accuracy, and the multi-layer perceptron of artificial neural networks was used as the implementation algorithm.

In order to improve the recognition rate, it was proposed to introduce additional filtering for the input signal of the neural network. In this way, the number of neurons in the hidden layer could be reduced and the recognition rate could be improved. Being able to reduce the number of neurons means increasing the likelihood that Neural network can be ported to smaller devices.

The further direction of this study is a more concrete version of this system by porting the functions of the Neural network circuit to the device.

References

A. Yun Young Bae. (2019). Music education for young children. Seoul: Dong Moon Sa. pp. 32-167
B. Cho Hong Gi. (1994). kodály's Principles in practice. Seoul: Book Publishing DaRa. pp. 23-126
C. Chae Wan Byeong., & Cho Hong Gi. (1999). kodály's Theory and basic education. Seoul: Book Publishing DaRa. pp. 24-86
D. Kodály, Zoltán. (1982). Folk music of Hungary Revised and Enlarges by Lajos Vargyas (3rd ed.). Budapest: Corvina Kiado. pp. 81-147
E. Kodály, Zoltán., & Ferenc, B. (Eds.). (1974). The Selected Writing of Zoltan Kodaly. (Lili, Hal´apy. Translator & Fred, Macnicol. Translator). London: Boosey & Hawkes. pp. 168-203
F. Lim Eun Young. (2013). A study on comparison of music education theory of Dalcroze, Kodaly, and Orff. Paper of Master's Degree, Ga chon University, Seongnam. pp. 35
G. Jeon hwang soo., & Kim hyun tak. (2020). "Trends in Motion Sensors Technology at Domestic and Foreign". Weekly ICT Trends, *1942*, pp. 16.
H. Jun Cheol Lee. (2018). "A Study on the Reliability and Validity of Measuring the Range of Motion of the Elbow Joint Using a Smartphone Application,". International Journal of Internet, Broadcasting and Communication(IJIBC), *Vol.10*, *No.3*, pp. 65-72.
I. Jang ju yong., Ryu moon ok., & Park soon chan. (2014). "Technology Trends of Range Image based Gesture Recognition,". 2014 Electronics and Telecommunications Trends, pp. 11.
J. Kwon Young Man., Kwon Yong Woo., Chung Dong Keun., & Lim Myung Jae. (2019). "The Comparison of Performance According to Initialization Methods of Deep Neural Network for Malware Dataset,". International Journal of Innovative Technology and Exploring Engineering (IJITEE), *Vol.8*, pp. 57-62
K. Gyöngyössy, N. M., Domonkos, M., Botzheim, J., & Korondi, P. (2019). "Supervised Learning with Small Training Set for Gesture Recognition by Spiking Neural Networks,". 2019 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 2201-2206