Development of Android-Based Speech Recognition Application Using Learning Vector Quantization Method in Optimizing Deaf Communication

Anggun Fergina
Nusa Putra University, Jln Raya Cibatu, Cisaat No.21, Sukabumi, West Java
anggun.fergina@nusaputra.ac.id

Adang Badrujaman
Nusa Putra University, Jln Raya Cibatu, Cisaat No.21, Sukabumi, West Java
adang.badru_ti18@nusaputra.ac.id

Indra Yustiana
Nusa Putra University, Jln Raya Cibatu, Cisaat No.21, Sukabumi, West Java
indra.yustiana@nusaputra.ac.id

ABSTRACT

Deafness is a condition of someone who has an abnormality in his hearing function. This condition can be temporary or permanent. There are two types of hearing loss that make a person deaf, the first is congenital deafness (present since birth), the second is deafness that occurs after birth. Congenital deafness is most likely caused by a genetic mutation, heredity from parents, or exposure to disease while still in the womb. While deafness that occurs after birth is usually caused by long-term reception of loud sound waves, age, injury, and is caused by certain diseases, such as infections of hearing. Communication is a message delivery activity that aims to establish a relationship between the communicant and the sender of the message. The problem that arises is how to communicate with deaf people, where communication is done using sign language. Whereas in the use of sign language, not everyone can and understands its use. So the researchers developed a speech recognition application with a speech to text feature to convert voice to text based on Android in making it easier to communicate with people with hearing impairment, so that everyone can communicate even without understanding certain sign languages. The Learning Vector Quantization method is a method for conducting learning in a supervised competitive layer, where the LVQ method is used to identify the accuracy and purity of the captured sound. The system development uses the Rapid Application Development (RAD) method.

Keywords
Speech recognition, speech to text, and learning vector quantization.

1. Introduction

Deafness is a condition of someone who has an abnormality in his hearing function. This condition can be temporary or permanent. There are two types of hearing loss that make a person deaf, the first is congenital deafness (present since birth), the second is deafness that occurs after birth. Congenital deafness is most likely caused by a genetic mutation, heredity from parents, or exposure to disease while still in the womb. While deafness that occurs after birth is usually caused by long-term reception of loud sound waves, age, injury, and is caused by certain diseases, such as infections of hearing (E. Juherna et al 2021).

WHO states that people with disabilities or disabilities in a country are about 10% of the total population in all. The Ministry of Health stated that the number of people with disabilities according to the results of the 2004 National Socio-Economic Survey (Susenas) was 6,047,008 people, consisting of 1,749,981 people with visual impairments (29%), 1,652,741 people with physical disabilities (27%), former sufferers of chronic disease 1,282,881 people (21%), mental retardation 777,761 people (12.8%), speech impaired 602,784 people (9.9%) [16]. From the source of data
received from the Head of Population of the Sirnajaya Village Government, there are 1.3% of the total population who suffer from hearing impairment.

Speech recognition or commonly known as automatic speech recognition (ASR) is a development of techniques and systems that allow computers to receive input in the form of spoken words (AB Jaman and A. Fergina 2021). This technology allows a device to recognize and understand spoken words by digitizing the words and matching the digital signal through a certain pattern stored in a device. The spoken words are then converted into digital signals by converting sound waves into a collection of numbers which are then translated with certain codes to identify the words.

Speech to text is a feature that can convert voice into written form. This allows computers to understand human language through voice commands (K. Nugroho 2019). In the process, the voice data used in speech to text is first converted to numeric data so that the computer can read it. The numerical data is then processed so that the computer can translate it into words. Speech to text is part of speech recognition, which is a field of computer and electronics science that deals with signals and systems, signal processing, signal enhancement, and others.

In general, speech recognizers process incoming voice signals and store them in digital form. The results of the digitization process are then converted in the form of a sound spectrum which will be analyzed by comparing it with the sound template in a database system (M. Siahaan 2020). To be able to achieve optimally for the system in processing the incoming voice signal, it is necessary to have a learning method in identifying the voice signal. Learning Vector Quantization (LVQ) is a network consisting of input units and output units with a Single Layer Feedforward architecture (Y. Aprizal, RI Zainal, and A. Afriyudi 2019). So that by applying this method, it is expected that the system can run optimally and function as expected.

2. Methodology

1. Field Study
   Is one way to obtain data by making direct observations of the object to be studied to obtain information about the problem.

2. Formulation of the problem
   In an effort to identify or find research problems, more than one problem is found. From these problems, it is necessary to choose one, which is the most appropriate and appropriate problem for research. If only one problem is found, the problem must also be considered whether or not it is appropriate and whether or not it is appropriate for research.

3. Study of literature
   A method used to collect data or sources related to the topic raised in a study. Literature studies can be obtained from various sources, journals, books, documentation, internet and libraries.

4. Data collection
   Data collection is in the form of:
   a. Primary Data is the source of research data obtained directly from the original source in the form of interviews, polls from individuals or groups (people) as well as the results of observations of an object, event or test result (object). In other words, researchers need to collect data by answering research questions (survey method) or object research (observation method).
   b. Secondary data is a source of research data obtained through intermediary media or indirectly in the form of books, records, existing evidence, or archives, both published and not published in general.

5. Data processing
   After the data is collected, then the data processing and analysis is carried out. Data analysis activities aim to give meaning and meaning to the data and are useful for solving problems in research that have been formulated.

6. Analysis
   After going through the data collection stage, the researcher must determine the type of analysis that will be used according to the level of research needs. Broadly speaking, data analysis is divided into two, namely:
   a. Non-statistical analysis
Non-statistical data analysis includes qualitative data, namely data that cannot be numbered, non-statistical analysis is more appropriate. Qualitative data is usually processed or analyzed based on its content (substance). This non-statistical analysis is often also called content analysis, which includes descriptive, critical, comparative, and synthesis analysis.

b. Statistical Data Analysis
Statistical data analysis includes quantitative data, namely data in the form of numbers or can be calculated, statistical analysis is more appropriate to use descriptive statistics and inferential statistics. Descriptive statistics are used to help describe (describe) the actual situation (facts) of a sample of investigations.

7. Conclusions and recommendations
At the end of a research, at the end there are always conclusions and suggestions. these two things were done after the researchers conducted analysis and interpretation, then the researchers made general conclusions (generalizations) based on the existing research boundaries and in accordance with the proposed hypothesis. Besides that, it is also necessary to provide suggestions, because research usually has limitations or assumptions.

a. Rapid Application Development (RAD) System Development Method
Rapid Application Development (RAD) system development method is a linear sequential software development process that emphasizes a short development cycle so that it can cut media development time faster (E. Hutabri 2019).

![Figure 1 Rapid Application Development (RAD) Method](image)

Stages The Rapid Application Development (RAD) system development method consists of:

1. Planning
   This stage is the initial stage in a system development, where at this stage identification of problems and collection of data obtained from users or user stakeholders is carried out which aims to identify the ultimate goal or purpose of the system and the desired information needs.

2. System Design
   In the system design stage, the activeness of the users involved is very important to achieve the goal because at this stage the design process and design improvement process are carried out repeatedly, if there are still design discrepancies to the user needs that have been identified in the previous stage. The output of this stage is a software specification that includes organization in the system in general, data structures, and others.

3. The process of developing and collecting feedback.
   At this stage, the system design that has been made and agreed upon is converted into a beta version of the application to the final version. At this stage, the programmer must continue to carry out development and integration activities with other parts by considering feedback from users. If the process runs smoothly, it can continue to the next stage, while if the application developed has not answered the needs, the programmer will return to the system design stage.

4. Product implementation or completion.
   This stage is the stage where the programmer applies the design of a system that has been approved in the previous stage. Before the system is implemented, the program testing process is carried out first to detect errors that exist in the developed system. At this stage, you can provide feedback on the system that has been made and get approval for the system.
b. **Learning Vector Quantization (LVQ) System Development Method**

In the research conducted, researchers created an application system that is able to capture sound waves and then convert them to writing or text with the **speech to text feature**. The system used is an artificial neural network (ANN) with a single-layer feed-forward network architecture type **consisting** of input units and output units. A **competitive layer** will automatically learn to classify input vectors. The classes obtained as a result of this competitive layer only depend on the distance between the input vectors. If the 2 input vectors are close to the same, then the competitive layer will put the two input vectors into the same class (Y. Aprizal et al. 2019).

![Figure 2 The learning vector quantization network architecture](image)

The steps of the **Learning Vector Quantization algorithm** consist of:

1. Initialization of initial weight (W) and LVQ parameters, namely maxEpoch, , dec and min
2. Input data input (X) and target class (T)
3. Set initial conditions: epoch = 0
4. Do if ; (epoch <maxEpoch) and ( mina )
   a. epoch = epoch + 1
   b. Determine J such that ||Xi-Wj|| at least use the Euclidian distance formula calculation
      \[ D(j) = (W_{ij} - x_i)^2 \]  
      (1)
   c. Fix Wj with the following conditions:
      If T = Cj then
      \[ W_j(t+1) = w_j(t)[x(t) - w_j] \]  
      (2)
      If T Cj then
      \[ W_j(t+1) = w_j(t)[x(t) - w_j(t)] \]  
      (3)
   d. Subtract the value of = * Deca
   Stop condition test with the output in the form of optimal weight

3. **Results and Discussion**

a. Needs Analysis
   1. **Functional Needs**
      Functional requirements are a requirement that must be owned by the system containing the processes or services provided by the system, where the system requirements include how the system must react to certain inputs and how the system behaves in certain situations. The functional requirements on the system built are:
      a. The system can detect sound signal as input.
      b. The system can convert voice to text.
      c. The system can automatically recognize Indonesian and English articulations.

   2. **Non-Functional Needs**
      In general, non-functional requirements have four kinds, including:
      a. Usability
         Namely non-functional requirements related to the ease of use of the system by the user.
      b. Portability
         Namely non-functional requirements that include the device or technology used to access the system. These devices or technologies are in the form of software, hardware, or network devices.
c. Reliability
   Namely non-functional requirements related to the reliability of the system and the security factor of the system.

d. Supportability
   Namely non-functional requirements where there is a support from the device for the use of the system being built.
   From the four non-functional requirements, it can be concluded that the system built has the following requirements:
   1. The system can be run on android devices.
   2. When the system is opened, will display a splash screen with a duration of 5 seconds.
   3. microphone button to start voice input.
   4. The system will automatically display the input results after the sound given stops for 2 seconds.
   5. The system must have permission to access the microphone on the android device used.

b. System planning
   1. Use Case Diagrams
      Use case diagram is a diagram that defines the design of the relationship between the user and the system. The use case of this application begins with the actor giving input in the form of a voice signal, the system receives the sound and then the voice signal is converted to written or text form as output. The use case diagram is shown in Figure 1.

![Use Case Diagram](image)

**Figure 3. Use Case Diagrams**

2. Use Case Scenario
   Use case scenario word records are in Table 1.

<table>
<thead>
<tr>
<th>Destination</th>
<th>Convert voice to text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor</td>
<td>User</td>
</tr>
<tr>
<td>Initial conditions</td>
<td>Action does not exist</td>
</tr>
<tr>
<td>Main Scenario</td>
<td>Voice input user</td>
</tr>
<tr>
<td>Scenario Alternative</td>
<td>Fail recording due to system error</td>
</tr>
<tr>
<td>Final Condition</td>
<td>The system captures the signal in the form of sound</td>
</tr>
</tbody>
</table>
3. Activity Diagrams

*Activity diagram* is a diagram that defines the activity flow diagram of the user and the application system, the activity will be separated by several tables that separate the user and the system. The *activity diagram* is depicted in Fig. **Figure 4**.

![Activity Diagrams](image)

**Figure 4. Activity Diagrams**

4. *Squance Diagrams*

*Squance Diagram* is a step or process of a system how an operation is carried out. The *activity diagram* is depicted in **Figure 5**.

![Squance Diagrams](image)

**Figure 5. Squance Diagrams**

c. Display Design

![Display Design](image)

**Figure 6. Display Design**
d. System Implementation
   The application system designed was then developed using Android studio software, where the speech to text feature was built for system needs.
   
   1. Splash Screen
      Splash Screen is the main display / window that appears when opening and running an application. The Splash Screen lasts 5 seconds to display a graphic consisting of a window containing an image, logo, and current software version.

   ![Figure 7. Splash Screen](image)

   Figure 7. Splash Screen

   2. Result menu
      Result menu is part of the application that is running to display the results of speech recognition, which is a series of text generated from the speech to text feature. In the main menu there is a microphone icon button that functions to start the process of capturing a sound signal when a voice is spoken, the maximum length of words that can be received by the system that is built is 35 words. After the voice input stops, the system will automatically display the text from the input results to the main page with a duration of 2 seconds.

   ![Figure 8. Result](image)

   Figure 8. Result

   e. System Test
      System testing in the accuracy of speech recognition of characters in the form of text generated from this application uses the Character Accuracy (Cacc) test [13]. To calculate the percentage of accuracy using the following formula:

      \[ cacc = \frac{\text{Jumlah kata yang dikenal dengan tepat}}{\text{Total semua kata yang di uji}} \times 100\% \]
Table 2. Text Test Results

<table>
<thead>
<tr>
<th>No</th>
<th>Test Sample</th>
<th>Results</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>I</td>
<td>In accordance</td>
</tr>
<tr>
<td>2</td>
<td>You</td>
<td>You</td>
<td>In accordance</td>
</tr>
<tr>
<td>3</td>
<td>Who with</td>
<td>Who with</td>
<td>In accordance</td>
</tr>
<tr>
<td>4</td>
<td>How are you</td>
<td>How are you</td>
<td>In accordance</td>
</tr>
<tr>
<td>5</td>
<td>Good afternoon</td>
<td>Good afternoon</td>
<td>In accordance</td>
</tr>
<tr>
<td>6</td>
<td>Healthy</td>
<td>Healthy</td>
<td>In accordance</td>
</tr>
<tr>
<td>7</td>
<td>Let us play</td>
<td>Let us play</td>
<td>In accordance</td>
</tr>
<tr>
<td>8</td>
<td>Please, sit</td>
<td>Please, sit</td>
<td>In accordance</td>
</tr>
<tr>
<td>9</td>
<td>Can I borrow</td>
<td>Can I borrow</td>
<td>In accordance</td>
</tr>
<tr>
<td>10</td>
<td>Do not disturb</td>
<td>Do not disturb</td>
<td>In accordance</td>
</tr>
<tr>
<td>11</td>
<td>Deaf</td>
<td>Deaf</td>
<td>In accordance</td>
</tr>
<tr>
<td>12</td>
<td>is</td>
<td>is</td>
<td>In accordance</td>
</tr>
<tr>
<td>13</td>
<td>state</td>
<td>state</td>
<td>In accordance</td>
</tr>
<tr>
<td>14</td>
<td>Hearing Function</td>
<td>Hearing Function</td>
<td>In accordance</td>
</tr>
<tr>
<td>15</td>
<td>Disturbance</td>
<td>Disturbance</td>
<td>In accordance</td>
</tr>
<tr>
<td>16</td>
<td>people</td>
<td>people</td>
<td>In accordance</td>
</tr>
<tr>
<td>17</td>
<td>Reason</td>
<td>Reason</td>
<td>In accordance</td>
</tr>
<tr>
<td>18</td>
<td>Disease</td>
<td>Disease</td>
<td>In accordance</td>
</tr>
</tbody>
</table>

From table 2 above, a system test was carried out where there were 18 words as a sound sample which were tested with the details of the words:

The number of Indonesian words = 18 words
Number of matching words = 18 words

\[ \text{Cacc} = \frac{18}{18} \times 100\% = 100\% \]

4. Conclusion

The conclusion from the results of the research that the author did was the development of the Speech application system Android-based recognition using the speech to text feature, where the system is built converts the voice signal as an input value into an output value in the form of text.

Development method system using the Rapid Application Development (RAD) method is very suitable in the cycle system development with a short duration of time so that it can cut time system development is faster. From the results of testing the accuracy of the system on the introduction of voice by converting it to text is 100% of the 18 words as a test sample.

References


MR Fadli, "Understanding the design of qualitative research methods," *Humanika.*, vol. 21, no. 1, pp. 33–54, 2021, doi:10.21831/hum.v21i1.38075.33-54


