Speaker-Dependent Based Speech Recognition

Lilik Untari[1] SF. Luthfie Arguby Purnomo[2] Nur Asiyah[3] Muhammad Zainal Muttaqien[4]
IAIN Surakarta
sastrainggrisiainska1@gmail.com

Abstract

This is the first part of the two parts of a qualitative focused R&D research aimed at designing an application to assist students with visual impairment (VI) in learning English writing and reading skills. The designed application was a speaker-dependent based speech recognition. Conducting alpha and beta testings, it was revealed that MAKTUM, the name of the application, exposed weaknesses on the selection of Ogden's Basic English as the linguistic resources for the application and on the recording complexities. On the other hand, MAKTUM displayed strengths in individualized pronunciation and simple interfaces to operate.

Key Words: *MAKTUM, Speech Recognition, Visual Impairment*

Abstrak

Penelitian pengembangan dengan fokus pada aspek kualitatif ini adalah penelitian tahap pertama dari dua tahap penelitian yang bertujuan untuk menghasilkan sebuah aplikasi berbasis *speech recognition* untuk membantu mahasiswa tunanetra dalam mempelajari bahasa Inggris khususnya keahlian menulis dan membaca. Setelah melakukan *alpha* dan *beta testing*, terungkap bahwa MAKTUM

[1] Teaching staff in English Letters Department IAIN Surakarta

[2] Teaching staff in English Letters Department IAIN Surakarta

[3] Teaching staff in English Letters Department IAIN Surakarta

[4] Teaching staff in English Letters Department IAIN Surakarta
They can be reached at sastrainggrisiainska1@gmail.com

memiliki kelemahan pada pemilihan *Basic English* oleh Ogden dan kompleksitas perekaman individual. Sementara itu kelebihan MAKTUM terlihat pada sistem pelafalan personal yang dimilikinya dan menu antar muka yang mudah dioperasikan.

Kata kunci: *MAKTUM, Speech Recognition, Mahasiswa Tunanetra.*

Introduction

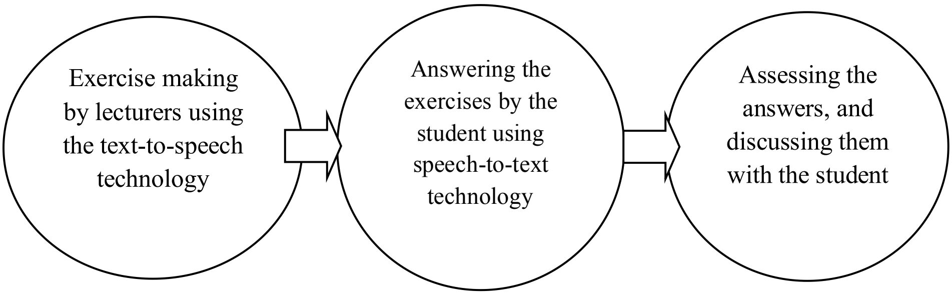
The absence of linguistics based learning aid for visually impaired (VI) students is one of the basic problems universities face and the same problem occurs in the Faculty of Islamic Studies and Teacher Training at State Islamic Institute of Surakarta (IAIN Surakarta). In its fourth year of inclusive education, the faculty static in developing any electronic or digital aid to sustain its visually impaired students in enhancing their learning experiences and achievements. In the scope of English Department student, one student with visual impairment suffers visual acuteness of 20/70 or classified as partial visual impairment (Berger and Constance, 1970). In WHO scale, 20/70 is classified into severe visual impairment (SVI) or low vision (Freeman, 2007).

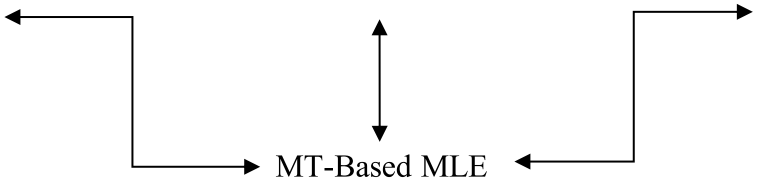
The PVI category the student suffers from, in the context of linguistics, triggers a high susceptibility toward the declination of linguistic proficiency (Galiano and Portelie, 2011). The linguistic profiencies the student suffers from are dominantly in reading and writing skills. This condition is perceptible from the necessity for the student to require a reading assistant when a test is in progress. In English language, abridging this condition, simplified English (SE) is designed (Kashdan and Barnes, 2002). This consideration to adopt simplified English is not yet taken into account when deciding to accept students with VI and thereby it triggers the feeling of social isolation (Webb, 2006), in the case of the student of English Department, the feeling emerges in reading and writing class. The social isolation is perceivable from the fact that the student is required to alter her reading and writing experiences into listening and speaking experiences exercised for reading and writing purposes. Therefore, an assistive technology to bridge these experiences is of necessity to assist the student in her reading and writing class. Text-to-speech technologies, like stylus pen (Zworykin and Flory, 1947), talking book pen, Kurzweil Reading Machine (KRM) (Dorman, 1995), The Reading Edge (Dorman, 1995) and Kindle. Meanwhile writing skills, the focus of this research, are abridgedable by speech-to-text technology. Microsoft Word, Dragon Nuance, Speakonia and other speech-to-text technologies are few to name.

The aforementioned speech recognition software and applications, due to its global nature, are operated based on speech following to the English pronunciation standards. This fact evokes a problem in the context of English as Foreign Language (EFL), a problem linked to the standardization of English pronunciation. The presence of English pronunciation standard on the speech recognition technologies indicates that the technologies are dominantly intended for English native speakers. This is problematic if connected to the problem faced by the VI student as aforementioned before. Students with VI possesses a misunderstanding and an incomplete comprehension of a sound (Wild, Wilson, and Hobson, 2013) from which language expressions of the students are limited especially in reading and writing skills. Therefore, speech-to-text technologies with standardized pronuciation are assumed to hinder non-native students, especially students with VI, when they attempt to learn reading and writing skills. Departing from this assumption and the fact for the need of speech-to-text technology abridgable for the VI student to use, this research and development inquiry with qualitative focus attempts to design speech-to-text technology friendly to VI users.

To design the technology, the first step taken was to adopt a concept of mother tongue related foreign language inspired from MT-Based MLE (Mother Tongue-Based Multilingual Education), a language instruction involving the use of mother tongue along with other languages used in a classroom (Malone, 2007). The incorporation of MT-Based MLE is expected to open a possibility to record voices and writings adjusted to the pronunciation standardization of the users. Thereby, it is expected the problems evoked from English pronunciation standardization in designing a speech recognition application are solved.

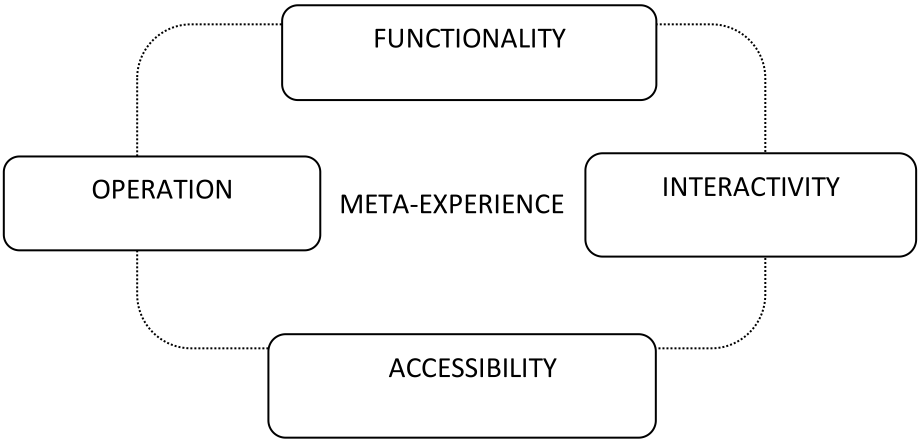
Departing from aforementioned logical sequence, this research focused on constructing a reciprocal design between student and lecturer. The following illustration might help explain the intended reciprocal design:


Exercise making by lecturers using the text-to-speech technology
Answering the exercises by the student using speech-to-text technology
Assessing the answers, and discussing them with the student



The reciprocality of the application is perceptible from the interactions between the lecturer and the student with the designed applicaition abridging them. The initiation of this reciprocality starts by the inputs executed by the lecturers to the student in writing skill exercises through text-to-speech technology. Utilizing this feature, the text composed by the lecturers is converted into a speech. Through the technology, the student performs an exercise by responding to the exercises given by the lecturer. The response is in the form of pseech by the student which is converted into a text by the application.

Completing the exercises, the exercise will be downloaded as a text to which the lecturers examine the answers and discuss the answers with the student. This cycle is expected to generate a meta experience, an experience resulting from thought and feelings toward the mood (Mayer and Gaschke, 1988), for the VI student. Regarding to writing skills for VI students, meta experience is visible from the writing engagement process executed through speech recognition technology. The following illustration depicts the relationship between the designed speech recognition application with the expected meta experience:



Opertion, functionality, interactivity, and accessibility are a union defining a program or an application. Operation refers to the mechanical capabilities a program or an application has. Functionality denotes the usabilities or benefits the program or the application has toward the users. Interactivity signifies the level and form of communication appearing between the users and the program or the application (HCI/Human Computer Interaction). Accessibility refers to the capabilities a program or an application has for an access by various types of users. Those four elements establish a meta-experience generated from the four elements the program or application has. For instance, the application this research attempts to design. The application aimed at assisting students with VI in learning English especially writing skill and thereby this application is expected to endow a real learning experience as that of non VI students. This application is expected to endow the users a situated and simulated learning, digital and virtual paedagogical presentation aligned to the real world (Shaffer, Squire, Halverson, Gee, 2004).

Therefore, it is an expectation that the students with VI possess meta-experience presented through situated and simulated learning generated from this English linguistics aid.

This research resulting in an application called MAKTUM focused, first, on writing skill as the primary focus of the skill and reading as the secondary focus. The primary focus of the object was the VI student and the secondary was the lecturer. Secondary focus emerged due to reciprocality owned by MAKTUM. Second, MAKTUM was a fusion of text-to-speech and speech-to-text with the former still under development. Third, MAKTUM was targeted for students with VI with categorization and clarification by Berger and Kautz of which the relationship between language acquisition and visual impairment is perceptible. Fourth, MAKTUM design related to speech recognition was limited to identification technology, sub technology linked to word recognition uttered by verification technology, sub technology designed to verify the uttered word accuracy. Fifth, MAKTUM did not incorporate NLG (Natural Language Generator), natural language verifier since MAKTUM utilized MT-Based MLE.

Research Objective

The purposes of this research were first, to reveal the weakness MAKTUM has in assisting writing skill learning by the VI student of English Department at IAIN Surakarta. Second was to unveil the strength MAKTUM has in assisting the writing skill learning, and third was to obtain responses from the VI student toward MAKTUM.

Research Methodology

This research was a research and development focusing not on product comparison but on response intakes toward the designed product in alpha and beta testing scope. The primary objective of this research was to design an application implemented to assist students with VI of English Department at State Islamic Institute of Surakarta in learning writing and reading skills.

This research was executed through three steps or triple helix (Mahdjoubi, 2009) namely preliminary research, design, and application. The preliminary research was aimed at revealing the negative impacts students with VI in the English Department without the presence of an assistive technology helping the students qualitatively. The result obtained from this preliminary research was utilized as an input to guide the researchers in designing MAKTUM. After the design was completed, the next phase was to acquire the responses toward MAKTUM from the student and the lecturers purposively.

The data validation utilized in this research was content-scale validation from Garcia-Valderrama and Mulero-Mendigorri. Content-scale validation specifically designed to validate research and development was executed through three phases namely selection, consultation, and scalation (2005). Selection revolves around qualitative and quantitative aspect selection assumed to possess the most crucial roles in developing a product. This selection was undergone by performing extensive reviews on related literature.

Consultation concerns on consultation to the experts regarding with the qualitative and quantitave aspects of the products. Scalation operates around employing scale design to imply the conjunction of qualitative and quantitative aspects. In the context of MAKTUM, qualitative aspects include interface and interactivity discussed through literary reviews discussing both aspects. The consultation of MAKTUM was executed by cooperating with I After Smile studio and the scalation was undergone by combining table structurization by Spradley intended to reveal the connection between interface and interactivity.

The research procedures were (1) observing the student with VI in English Department regarding with the level of VI and English writing and reading competences (2) interviewing the student regarding with the difficulties faced in English writing and reading skills (3) designing the alpha version of MAKTUM based on the obsevation and the interview (4) performing an alpha testing on MAKTUM involving the experts from I After Smile Studio (5) redesigning MAKTUM based on the result of alpha testing (6) performing a beta testing involving the student and the lecturers of reading and writing (7) uploading to velis.xyz and Google Play.

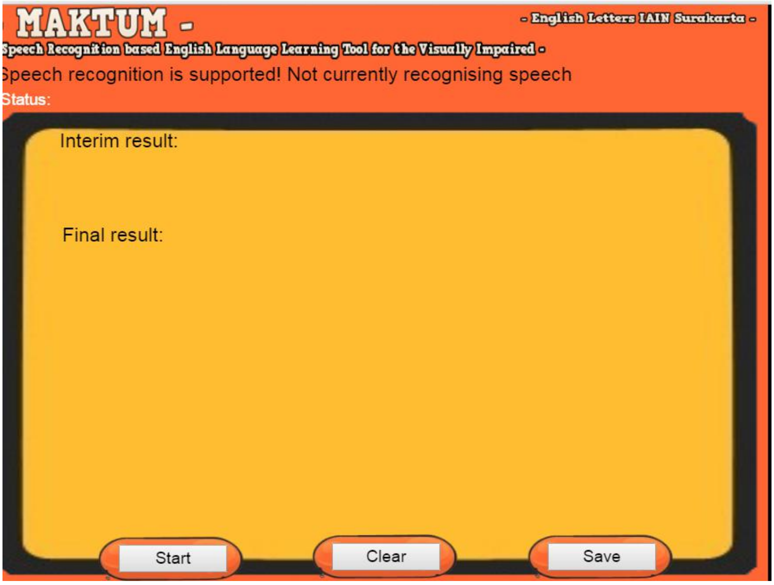
The Description of MAKTUM

MAKTUM yang bisa diunduh untuk android melalui https://play.google.com/store/apps/details?id=com.iaftersmile.maktum atau bisa diakses online melalui https://dl.dropboxusercontent.com/u/110823528/Maktum%20x2/index.

html ini didesain dengan menggunakan pendekatan koneksionis dan karakterisasi ASR (*Automatic Speech Recognition*) (Boulard and Morgan, 2012) untuk fitur *speech recognition*-nya dan pendekatan pola (*pattern*) untuk menu antar mukanya. Pendekatan koneksionis yang digabungkan dengan ASR menghasilkan aplikasi *speech recognition* yang bersifat *speaker-dependent*, berkosa kata khusus, dalam kasus ini adalah *Basic English*, dan tuturannya bersifat isolatif.

MAKTUM is downloadable for Android application via https://play.google.com/store/apps/details?id=com.iaftersmile.maktum or accessible via online from https://dl.dropboxusercontent.com/u/110823528/Maktum%20x2/index.html was designed using a connectionist approach and characterization of ASR (Automatic Speech Recognition) for its speech recognition feature, and pattern approach for the inter-face menu. Connectionist approach combined with ASR produced speaker-dependent speech recognition application having special vocabularies, in this case Basic English, and isolative speech.

The resulting design from the aforementioned approach and characterization is a reflection of the special linguistic needs for students with visual impairment. The underlying assumption is that those who have disabilities, although inclusive in domain, require special assistance also both in terms of operationalization and content of a linguistic aid application. This assumption is in line with the concept of HCI (Human Computer Interaction) in the perspective of the pattern approach, which stresses the specificity of the needs of the technology users (Borchers, 2001). In the context of an application intended for users with special needs, the interface of MAKTUM was designed with a minimalistic number of menu and simple functionality. The visually impaired student from whom MAKTUM was designed has low vision so that the menu design applies bright colors in order to be easy to read. The example of menu presented in MAKTUM is as follows:


MAKTUM -

- English Letters IAIN Surakarta -

Speech Recognition based English Language Learning Tool for the Visually Impaired -

Speech recognition is supported! Not currently recognising speech

Status:

Interim result:

Final result:

Chart 1 The Display of MAKTUM

The combination of connectionist approach, ASR characteristics, and pattern approach in the context of linguistic need for those with special needs produces in the application that is in linearity with the user, or in simple words, user-friendly.

MAKTUM consists of three main menus, namely Start, Clear, and Save. The Start Menu is to start the recognition, CLEAR is to delete the text resulted from the recognition, and Save is to save and download the text. In addition to the operational menu, there are also descriptive menus which describe the status of the spoken utterances; Interim Result and Final Result menus. Interim Result Menu displays the text version of the utterances that can still be corrected while Final result Menu displays the final result of the utterances.

Alpha and Beta Testing Result

The Alpha and Beta Testing implemented in this research covers the alpha testing for content and operationalization as suggested in the concept of usability by Craig and Jaskiel (2002). Through the Alpha Testing from the language application design expert in I After Smile studio, MAKTUM is claimed to have two strengths and two weaknesses. The first strength is that MAKTUM incorporates individual-based pronunciation which minimizes the basic problem in pronunciation, International standard pronunciation. This strength, however, comes with a weakness, the complexity of the recording.

The complexity comes to appear as the user has to record around 1500 words covered in Charles Kay Ogaden's theory of Basic English. From linguistic perspective, the decision to use this Basic English is considered less appropriate as this theory, in addition to the fact that it is too classic to use, does not have the linguistic and philosophical accuracy in defining the meaning of Basic and English as disclosed by Flesch (1944), and the trend of free to speech in syntactical context which in turn lead to grammatical confusion for Basic English has different grammar from the English language in general (1994).

The linguistic weakness of Basic English, as mentioned previously, is based more on the structural perspective. From the functional perspective in the context of disabilities, as discussed by oleh Pena (1967), Becker (1977), Templer (2006), dan Templer (2009), it is revealed that English simplification is needed for those with special needs. The problems that arise in the context of language simplification lies in the question whether or not the simplification of the language includes all language units or only one of them such as phonetic (Santa Ana, 1991) or lexica; (Spacia, Jauhar, Mihalcea, 2012) simplification.

The decision to choose Ogden's Basic English is in relation to the holistification of English simplification which covers not only word-based lexical simplification with high frequency of use but also the restructuring of English grammar though the implementation in MAKTUM becomes redundant. The redundancy comes to appear as the grammatical structure does not become the main focus in its design. MAKTUM which function as a speech recognition based linguistic aid application has the purpose of transferring the user's voice into a physical lexical form so that third parties such as lecturers or teachers can, in the context of writing or grammar lectures, read the work of MAKTUM user, in this case the student.

The basic function, as described previously, does not require grammatical construction as formulated by Ogden. The separation of frequent vocabularies from the grammar in MAKTUM is because the design of MAKTUM was based on the connectionist approach, ASR characteristics, and pattern approach which emphasize on specific design for specific user.

The second strength, as revealed from the Alpha and Beta Testing, in the context of the user interface of the speech recognition designed based on pattern approach is that MAKTUM has simple user interface resulting in the easy operation by the user. In the perspective of speech recognition with speaker-dependent characteristic, the simple user interface menu designed based on the specificity of the user has fulfilled the criteria of being user-friendly. However, in the perspective of speaker-dependent speech-recognition and the perspective of users with special needs in overall context, the user interface displayed in MAKTUM does not meet the criteria of being user-friendly.

This claim is based on the perspective of correlation of the menus at the user interface categorized as collection, a relationship among objects (menus) directly related to the operationalization of a menu without affecting other objects in the interface (Gallitz, 2007). In this perspective of collection, the expert team in I After Smile consider, MAKTUM should maximize the user interface menu for all levels of the visually impaired people by implementing voice-recognition or motion sensors to the interface menu operationalization which basically emphasizes on the ergonomics of menu.

Conclusion

By means of applying the Alpha Testing from the language application design expert in I After Smile studio, MAKTUM is claimed to have two strengths and two weaknesses. The first strength is that MAKTUM incorporates individual-based pronunciation which minimizes the basic problem in pronunciation, International standard pronunciation. This strength, however, comes with a weakness, the complexity of the recording.

References

American Libraries, Vol. 26, No. 11 (Dec., 1995), pp. 1143-1144.

Becker, W. (1977). Teaching reading and language to the disadvantaged-What we have learned from field research. *Harvard Educational Review*, *47*(4), 518-543.

Berger, Allen dan Constance R. Kautz. (1970). *Sources of Information and Materials for Blind and Visually Limited Pupils. Elementary English, Vol. 47, No. 8 (December, 1970), pp. 1097-1105.*

Borchers, J. O. (2001). A pattern approach to interaction design. *Ai & Society*,*15*(4), 359-376.

Bourlard, H. A., & Morgan, N. (2012). *Connectionist speech recognition: a hybrid approach* (Vol. 247). Springer Science & Business Media.

Craig, R. D., & Jaskiel, S. P. (2002). *Systematic software testing*. Artech House.

Dorman, David. (1995). *Technically Speaking: Products for the Blind and Visually Impaired.*

Flesch, R. (1944). How Basic is Basic English?. *Harper's Magazine*, *188*(1126), 339-343.

Freeman et. al. (2007). *Care of the Patient with Visual Impairment (Low Vision Rehabilitation*). American Optometric Association.

Galiano, Anna. R dan Serge Portalier. (2011). *Language and Visual Impairment: Literature Review.* International Psychology: Practice and Research. Vol. 2.

Galitz, W. O. (2007). *The essential guide to user interface design: an introduction to GUI design principles and techniques*. John Wiley & Sons.

Garcia-Valderrama, Teresa dan Eva Mulero-Mendigorri. (2005). R&D Management 35, 3, 2005. Blackwell Publishing.

Kashdan, Sylvie dan Robby Barnes. (2002). *Teaching English as a New Language to Visually Impaired and Blind ESL Students: Problems and Possibilities. Kaizen Program for New English Learners with Visual Limitations.*

Mahdjoubi, Darius. (2009). *Four Types of R&D*. Presentation Paper. University of St. Edward. Texas

Malinowski, B. (1994). The problem of meaning in primitive languages.*Language and literacy in social practice: A reader*, 1-10.

Malone, Susan. (2007). *Mother Tongue-Based Multilingual Education: Implications for Education Policy.* Proceeding of he Seminar on Education Policy and the Right to Education: Towards More Equitable Outcomes for South Asia's Children Kathmandu, 17-20 September 2007

Mayer, D. John dan Yvonne N. Gaschke. (1988). *the Experience and Meta-Experience of Mood*. Journal of Personality and Social Psychology Vol. 55, No. 1, 102-111.

Pena, A. A. (1967). A Comparative Study of Selected Syntactical Structures of the Oral Language Status in Spanish and English of Disadvantaged First-Grade Spanish Speaking Children.

Santa Ana, A. (1991). Phonetic simplification processes in the* English of the Barrio: Across-generational sociolinguistic study of the Chicanos of Los Angeles.

Shaffer,W.D., Squire, R.K., Halverson, R., Gee. P.J. (2004). *Video Games and Future Learning.* Academic Advanced Distributed Learning Co-Laboratory. University of Wisconsin Madison: US.

Specia, L., Jauhar, S. K., & Mihalcea, R. (2012, June). Semeval-2012 task 1: English lexical simplification. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation* (pp. 347-355). Association for Computational Linguistics.

Templer, B. (2006). Revitalizing'Basic English'in Asia: New directions in English as a lingua franca. *TESL Reporter*, *39*(2), 17-33.

Templer, B. (2009). A two-tier model for a more simplified and sustainable English as an international language. *Journal for Critical Education Policy Studies*, *7*(2), 187-216.

Webb, Sue. (2006). *Can ICT Reduce Social Exclusion? The Case of an Adults' English Learning Programme*. British Educational Research Journal Vol. 32, No. 3, June 2006, pp.481-507.

Wild, T. A., Hilson, M. P., & Hobson, S. M. (2013). *The Conceptual Understanding of Sound by Students with Visual Impairments*. Journal Of Visual Impairment & Blindness, 107(2), 107-116

Zworykin, V. K. dan L.E. Flory. (1947). *an Electronic Reading Aid for the Blind.* Proceedings of the American Philosophical Society, Vol. 91, No. 2 (Apr. 5, 1947), pp. 139-142.