SPEECHPAL: SPECIALIZED SPEECH AID DEVICE FOR THERAPISTS ASSISTING CHILDREN WITH REPAIRED CLEFT LIP AND PALATE

Jafeth C. Estocado
Jan Carlo A. Magpantay
Joyce Ann P. Precilla

College of Engineering
First Asia Institute of Technology and Humanities, Tanauan City, Batangas, Philippines
jafethestocado@gmail.com, jncrlmgpnty020@gmail.com, joyceannprecilla29@gmail.com

ABSTRACT

Children with repaired cleft lip and palate (CLAP) often face challenges in speech clarity and communication, necessitating tailored therapeutic support. In response, the researchers designed and implemented SpeechPal, an innovative assistive technology intended to aid speech therapists in providing effective therapy. The system integrates three key features: text-to-speech and speech-to-text conversion, nasal airflow detection, and augmentative and alternative communication (AAC) functionality. The project implementation involved Intel Core i7-9700 (9th Gen) with MSI GeForce GTX 1050 Ti Dual Fan OC, Ypa 4016 Headset Microphone, HXV710 Nasal Air Flow Sensor, ANMITE 14" HDR IPS FHD Portable LED Gaming Touch Monitor, HuBERT Model, Google Speech Recognition, and PyQt6-Based Graphical User Interface for AAC. The results showed that the researchers successfully implemented the system and achieved the objectives of the study. This research highlights the potential of SpeechPal to address speech therapy challenges for children with repaired CLAP, offering a reliable, efficient, and user friendly solution that promotes inclusivity and improved communication.

1. 0 INTRODUCTION

Cleft lip and palate (CLAP) is one of the most common birth defects in the Philippines, affecting about 1 in every 1,000 live births. This congenital condition results from improper fusion of the lip and palate during early pregnancy and can cause significant speech and communication challenges. Dental anomalies like missing or misaligned teeth impair articulation, while velopharyngeal insufficiency—affecting a third of patients post-surgery—disrupts airflow during speech, causing hypernasality and unclear speech[1]. Hearing impairments from defective Eustachian tubes further impact communication and quality of life.

Despite surgical repair between three to nine months of age, many children continue to struggle with speech clarity. Traditional speech therapy relies on subjective assessments and lacks real-time feedback, making it hard for pathologists to monitor progress. This study introduces SpeechPal, an assistive therapy device that uses advanced technologies to improve speech clarity and communication in children with repaired CLAP.

SpeechPal features nasal air emission (NAE) detection, speech-to-text translation, and augmentative and alternative communication (AAC). NAE detection helps monitor excess nasal airflow, critical for children with velopharyngeal insufficiency. AAC offers alternative communication methods for those with articulation issues, supporting social interaction and reducing frustration. Speech-to-speech translation enhances communication by converting unclear speech into a more understandable format.

The design also considers engineering constraints to meet technical, ethical, and social standards, prioritizing inclusivity, data privacy, and therapeutic integrity. The study, in collaboration with Early Start Therapy Center in Tanauan, will run for six and a half months and involve children aged 3 to 10, speech therapists, and caregivers, aligning with key developmental milestones. By integrating NAE detection, AAC, and speech-to-speech translation, SpeechPal aims to improve therapy effectiveness and empower children to communicate more clearly.

1.1 Objectives of the Project

The study aims to develop a specialized speech aid device that provides therapists with a tool to support personalized and effective speech therapy interventions for children with repaired cleft lip and palate. By the end of the research, the following specific objectives will be obtained:

- To evaluate different existing models for speech recognition tailored for children with repaired cleft lip and palate.
- To design a user-friendly graphical interface for SpeechPal that facilitates ease of use for therapists during personalized speech therapy sessions.

- To evaluate SpeechPal's impact on speech therapy outcomes at the Early Start Therapy Center in Tanauan.
- To gather and analyze therapist feedback to assess the performance efficiency and usability of SpeechPal during therapy sessions.

1.1.1 Scope and Limitations of the Project

This study focuses on the development of SpeechPal, a speech recognition system designed as a therapy aid for children with repaired cleft lip and palate (CLAP). The system is intended solely for therapeutic applications and is not recommended for daily communication, serving as a supplementary tool rather than a substitute for professional therapy. SpeechPal is specifically tailored to process speech from individuals with CLAP and may not provide accurate recognition for users with other speech disorders, such as Down syndrome or aphasia. The system incorporates a nasal airflow detection feature, which requires precise positioning of the sensor to detect airflow directly from the nose; however, it does not measure the exact volume of nasal airflow. Additionally, SpeechPal includes a built-in word correction system for the speech-to-text component, which is limited to the English language. Environmental conditions, such as echo and reverberation, can impact the system's accuracy and reliability, with optimal performance achieved in acoustically controlled environments. Furthermore, the system is designed for children aged three to ten years, and its applicability beyond this age range has not been validated.

2. 0 REVIEW OF RELATED WORK

Cleft lip and palate (CLAP) is a congenital condition that affects the upper lip and/or palate, leading to significant speech disorders. The severity of the cleft directly influences speech development, with children having more severe forms of clefts facing challenges in producing speech sounds. Cincinnati Children's Hospital emphasizes that comprehensive care, including surgery, speech therapy, and technological interventions, is essential for improving communication skills and reducing the impact of CLAP on affected children[2]. Therapy is particularly crucial for addressing speech disorders such as nasal air emission and hypernasality, which are common in children with CLAP.

In speech therapy, speech recognition systems have emerged as valuable tools for enhancing therapy outcomes. Speech recognition technologies, including augmentative and alternative communication (AAC) devices, provide real-time feedback to children during therapy sessions. This feedback is essential in helping children with CLAP improve speech clarity and fluency. Such technologies allow therapists to track a child's progress and personalize interventions, ultimately making speech therapy more engaging and effective for children with complex speech needs.

The role of technology in speech therapy has grown significantly, particularly for children with CLAP who face challenges accessing specialized care. Najm et al. highlight how robots, like Robot Lily, can provide interactive platforms that simulate human-like contact, offering speech feedback and motivation during therapy[3]. These robots not only enhance engagement but also help bridge the gap in therapy accessibility, particularly for children in underserved areas. While there are challenges in replicating human-like interactions, the potential of robotics to transform speech therapy for children with CLAP is undeniable, offering unique alternatives for therapeutic interventions.

The effectiveness of technological interventions in speech therapy has been underscored by their role during the COVID-19 pandemic when traditional face-to-face sessions were disrupted. Wiśniewska observes that ICT tools, such as video conferencing and interactive applications, helped maintain continuity in speech therapy sessions, proving especially beneficial for students who were initially skeptical of using digital tools[4]. These tools allow therapists to provide personalized learning experiences, enhancing therapy engagement and ensuring that children receive consistent support even when in-person sessions are not feasible.

Functional words are critical in augmentative and alternative communication (AAC) systems, particularly for children with CLAP, as they facilitate meaningful and functional communication. Core vocabulary words, such as "go," "want," and "help," are essential for enabling children to express a wide range of thoughts and needs. These words form the foundation of AAC communication, promoting language development and helping children construct sentences in different contexts. Using core vocabulary to model simple phrases enhances children's ability to engage in meaningful conversations, boosting their confidence in expressing themselves.

3.0 METHODOLOGY

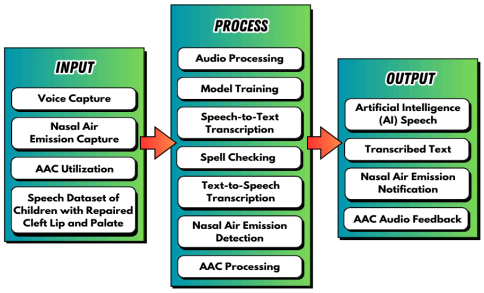

Fig. 1. Conceptual Framework

The system's IPO framework, shown on Fig.1, centers on patient voice capture, nasal air emission detection, AAC integration, and a speech dataset for children with cleft lip and palate. Voice capture uses microphones and PyAudio for real-time speech recording, while airflow sensors and signal processing detect nasal emissions. AAC is integrated through custom software with visual interfaces and audio feedback, allowing users to interact via categories and images. A speech dataset of children aged 3–10 supports system functionality.

The system processes three main features: speech translation, nasal emission detection, and AAC. Audio is pre-processed using Librosa, with AI models built on TensorFlow and PyTorch handling speech and emission analysis. Speech-to-text uses Google's API, followed by word checking and Google TTS for audio output. Nasal data is analyzed for airflow patterns, while AAC inputs are handled through Python's PyQt6 for visual and auditory responses.

Outputs include transcribed text, nasal emission indicators on-screen, and AAC audio feedback, offering real-time analysis and interactive support to improve therapy for children with cleft conditions.

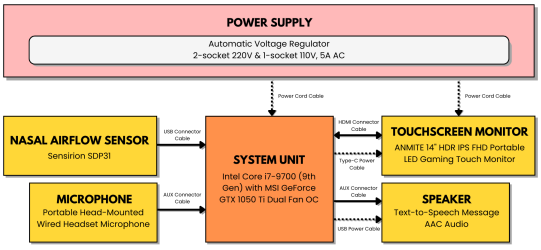

Fig. 2. Block Diagram

Fig. 2 illustrates the placement of hardware components in the device, showcasing the interaction between the system unit and the input/output components. At the core is the Intel Core i7-9700 (9th Gen) processor, housed within the system unit alongside an MSI GeForce GTX 1050 Ti Dual Fan OC, serving as the command center of the device. Power for the setup is generated through a voltage regulator automatic circuit, capable of operating at 220V and 110V AC with a 5A current rating to ensure proper functionality of all components. The system is integrated with a nasal airflow sensor (HX710) via a USB connector cable, enabling the detection of nasal airflow, while a moveable, head-worn wired headset microphone connects through an AUX cable to capture audio input. For output, the system unit interfaces with a 14-inch ANMITE HDR IPS FHD portable LED gaming touch monitor using an HDMI connector and Type-C power cable, providing a touchscreen display for user engagement. Additionally, audio output is delivered through a speaker connected via an AUX cable, facilitating Text-to-Speech messages and AAC audio. Both the monitor and speaker draw power directly from the power supply, ensuring consistent performance. This setup seamlessly integrates various peripherals, with the system unit serving as the hub for all input and output processes.

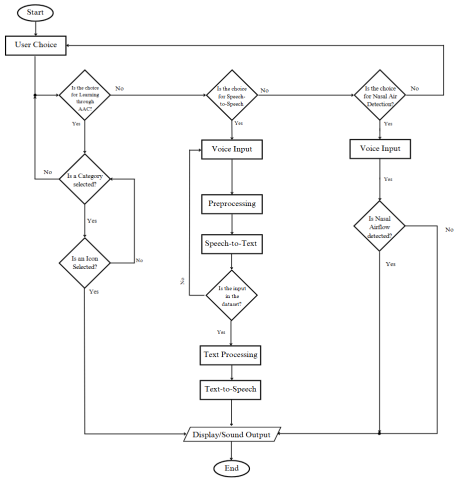

Fig. 3. Flow Chart

In Fig. 3, the flowchart outlines the system's decision-making and processing flow for three user options: Learning through AAC, Speech-to-Speech, and Nasal Air Detection. For Learning through AAC, the system verifies category and icon selection before proceeding to display or sound output. In Speech-to-Speech, voice input undergoes preprocessing, speech-to-text conversion, dataset validation, text processing, and text-to-speech feedback before delivering the final output. For Nasal Air Detection, voice input is analyzed for nasal airflow, with results displayed or outputted as sound. This structured flow ensures efficient user interaction, tailored outputs, and real-time assistance based on the chosen mode of operation.

4.0 RESULTS AND DISCUSSION

Fig. 4. Prototype Setup (Hardware)

The SpeechPal prototype is a 28-inch-tall kiosk designed for seated speech therapy sessions, providing a formal yet comfortable environment as depicted on Fig.4. It features a 14-inch touchscreen monitor, a nasal airflow sensor, a speaker, and a microphone, all powered by an integrated system unit. This setup ensures ease of use and accessibility for children with repaired cleft lip and palate, guided by therapists during sessions.

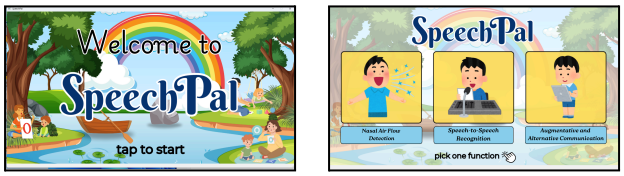

Fig. 5. Preliminary Page and Main Dashboard

The system application includes a preliminary page and a main dashboard for feature navigation. The preliminary page provides an introductory interface with an overview of the software's purpose, creating a user-friendly starting point, as shown in Fig 5. From there, users access the main dashboard, which features Nasal Airflow Detection, Speech Recognition, and Augmentative and Alternative Communication (AAC) for seamless feature selection

Table 1. Summary of Evaluations of the Accuracy of Speech Recognition Models

Model	Evaluation Metric		
	Word Error Rate	Latency	Real-Time Factor
Google Speech-to-Speech	32%	1.397 s	0.292
Wav2Vec 2.0	100%	2.102 s	0.449
HuBERT	18%	2.11 s	0.38

The SpeechPal device underwent validation testing using three speech recognition models—Google Speech-to-Text, Wav2Vec 2.0, and HuBERT—evaluated on Word Error Rate (WER), latency, and Real-Time Factor (RTF) using ten recorded phrases of a child with a repaired cleft lip and palate. As summarized in Table 1, the Google Speech-to-Text model exhibited a WER of 32%, latency of 1.397 seconds, and an RTF of 0.292, demonstrating fast processing speed but limited adaptability due to its inability to fine-tune for specialized speech patterns. Wav2Vec 2.0 showed a WER of 100%, latency of 2.102 seconds, and an RTF of 0.449, but its failure to transcribe speech accurately for this demographic made it ineffective despite fine-tuning capabilities. HuBERT outperformed the other models with a WER of 18%, latency of 2.11 seconds, and an RTF of 0.38, demonstrating accurate transcriptions with its ability to be fine-tuned on speech recordings from children with cleft lip and palate, making it the most suitable model for integration into the SpeechPal device.

Fig. 6. Nasal Air Flow Detection

When the Nasal Airflow Detection feature is activated, the system identifies the presence of nasal airflow and provides real-time feedback. Visual indicators on the display, as shown in the Fig 6, enable users to easily determine whether airflow has been detected. This immediate feedback supports efficient and accurate interpretation during speech therapy sessions.

Fig. 7. Speech-to-Speech Recognition

The Speech Recognition feature, depicted in Figure 7, processes spoken language and converts it into text, accommodating the unique speech patterns of children with repaired CLAP. This ensures accurate transcription while offering valuable support for therapists during sessions. By providing real-time feedback, the feature aids in overcoming challenges in understanding specific spoken words.

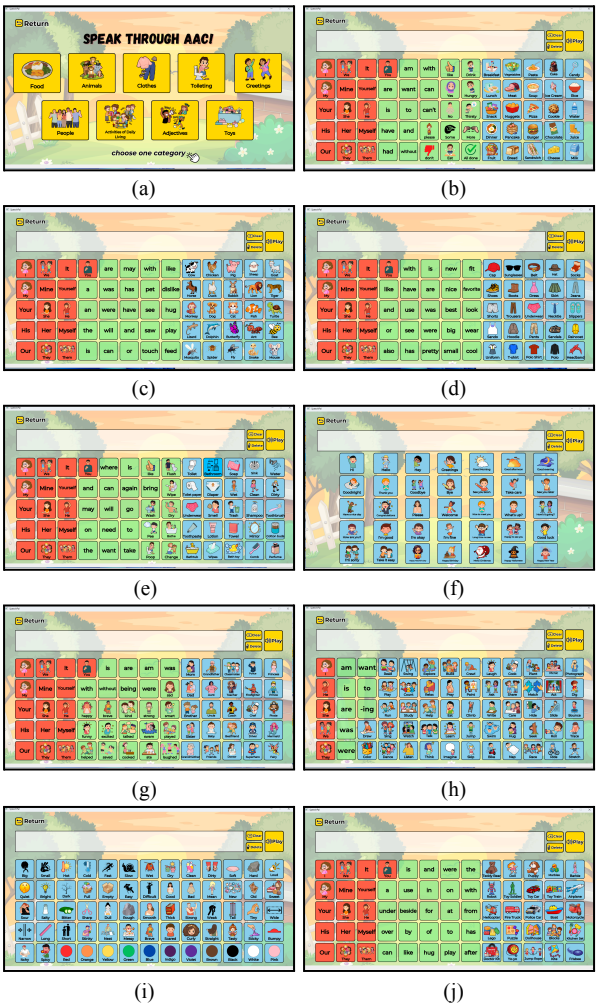

Fig. 8. AAC Categories

The Augmentative and Alternative Communication (AAC) system, shown in Fig. 8(a), includes nine communication categories designed for children aged 3–10. These categories—such as food, animals, greetings, and toys—are illustrated in Figs 8(b) to 9(j) and offer an intuitive interface using images and words to support effective, user-friendly communication.

The project was implemented at the Early Start Therapy Center in Tanauan, Batangas, to assess SpeechPal's impact on speech therapy outcomes. Three children with repaired cleft lip and palate participated, with evaluations conducted by speech pathologists using criteria from the CAPS-A (2006) and UPS (2016) frameworks. A pre-test established a baseline of the participants' speech conditions, while a post-test after using SpeechPal measured improvements, allowing for an evaluation of the device's effectiveness in addressing speech-related issues by comparing the rate of improvement in child's speech.

Table 2. Comparison of Average Pre-Test and Post-Test Scores for Oct-Nov and Nov-Dec Interventions

Children No.	October - November Intervention (Without SpeechPal)		November-December Intervention (With SpeechPal)	
	Pre-Test	***Post-Test***	***Pre-Test***	***Post-Test***
1	2.5	2.5	2.5	1.5
2	2.33	2.17	2.17	1.17
3	2.33	2.17	2.17	1.17

Table 2 compared traditional speech therapy to therapy incorporating SpeechPal for three children with repaired cleft lip and palate. Traditional therapy (October-November) showed minimal improvement, with Child 1 remaining at 2.5 and Children 2 and 3 slightly improving from 2.33 to 2.17. Adding SpeechPal (November-December) led to significant progress, with Child 1 improving to 1.5 and Children 2 and 3 advancing to 1.17, highlighting the effectiveness of real-time feedback.

The feedback evaluation assesses SpeechPal's usability and performance efficiency based on ISO 25010 Software Quality Standards. Approved by the thesis adviser and a speech pathologist, the form ensures validity. Expert feedback helps identify strengths and areas for improvement, supporting the system's ongoing enhancement.

Table 3. Results of Performance Efficiency Evaluation

	Criteria		
	1	**2**	**3**
Average Score	4	4	4

Shown in Table 3 are the average scores given by three speech pathologists, with all criteria under performance efficiency receiving a consistent average score of 4, indicating strong agreement on the system's effectiveness.

Table 4. Results of Usability Evaluation

	Criteria					
	1	**2**	**3**	**4**	**5**	**6**
Average Score	4	4	4	4	4	4

Presented in Table 4 shows that all three speech pathologists gave a consistent score of 4 across all usability criteria, resulting in an overall average usability score of 4, reflecting uniform positive feedback.

5.0 CONCLUSION

The completion of SpeechPal, a specialized speech aid device for therapists assisting children with repaired cleft lip and palate, led to several conclusions. The researchers evaluated speech recognition models and found HuBERT to be the most effective, with a word error rate of 18%, latency of 2.11 seconds, and a real-time factor of 0.38, ensuring reliable performance for therapy applications. A user-friendly graphical interface (GUI) was successfully developed using PyQt6, allowing therapists to navigate the system efficiently and conduct personalized therapy sessions. Implemented at the Early Start Therapy Center in Tanauan City, the project included data collection, demonstrations for therapists, system familiarization, and prototype testing. The evaluation at the center demonstrated SpeechPal's effectiveness in improving therapy outcomes, with significant progress observed during the November–December intervention period using SpeechPal compared to the October–November period without it. Real-time feedback from SpeechPal led to enhanced speech clarity, reduced nasal resonance, and better airflow control, affirming its value in speech therapy. Therapist feedback confirmed the device's performance efficiency and usability, with both metrics scoring an average of 4.0, highlighting its prompt processing, reliable AAC interactions, consistent performance, and intuitive interface.

6.0 RECOMMENDATIONS

Following the completion of SpeechPal, several recommendations were proposed to enhance its functionality and impact. Portability should be prioritized to facilitate use across various therapy settings, while expanding the study's scope to include diverse participants could broaden its applicability. Advanced speech recognition models should be explored for improved transcription accuracy, and incorporating additional audio recordings, particularly in Filipino, would enhance local context relevance. More sensitive nasal airflow sensors could improve detection accuracy, and the AAC functionality should be enhanced with new categories and an expanded vocabulary. A login feature could enable progress tracking, while making the software cloud-based would provide online accessibility and centralized data management. Multilingual capabilities should be extended to meet diverse linguistic needs, and gamification could engage children more effectively during therapy. Adaptive machine learning models could personalize therapy, and adapting SpeechPal for home use would ensure continuous support outside therapy sessions. Finally, promoting the device in government and institutional settings, as well as collaborating with therapy centers, would validate its effectiveness in diverse environments and extend its reach.

7.0 ACKNOWLEDGMENT

The researchers extend their heartfelt gratitude to all who contributed to the success of this study. First and foremost, sincere thanks to God for His guidance and blessings throughout this journey. Special appreciation is given to Engr. Marco A. Burdeos, thesis adviser, and Engr. Francis A. Malabanan, capstone project professor, for their invaluable guidance and support. Gratitude is also extended to the panelists—Engr. Adonis S. Santos, Engr. Sherryl M. Gevaña, and Engr. Favis Joseph C. Balinado—for their expert insights. The researchers thank Early Start Therapy Center Tanauan for their cooperation and participation, as well as Xinyx Design Consultancy and Services Inc. and the City Government of Tanauan for their financial support. Deep appreciation goes to the 4th-year Electronics and Computer Engineering students, as well as to families and friends, for their unwavering encouragement. Special thanks to the families who participated in speech recording sessions and to the Magpantay family for their hospitality and support.

8.0 REFERENCES

1. Kummer, A. W. (n.d.). Talking about … Cleft and Palate: The Effects on Communication Skills. Cleft Palate and Craniofacial Anomalies: The Effects on Speech and Resonance, 3rd Edition. https://www.smiletrain.org/sites/default/files/2021-01/cleft-palate-effects-on-communication.pdf

2. Children's Hospital Los Angeles. (n.d.). Children's Hospital Los Angeles. https://www.chla.org/cleft-lip-and-palate

3. Najm, A., Chew, E., & Bentley, B. (2023). Robot-Assisted Language Education and Speech Therapy for Children with Cleft Lip and Palate. Proceedings of the . . . European Conference on E-learning, 22(1), 202–211. https://doi.org/10.34190/ecel.22.1.1787

4. Wiśniewska, J. (2020). Speech therapy students' attitudes to the use of ICTs in speech therapy practice. Interdyscyplinarne Konteksty Pedagogiki Specjalnej/Interdisciplinary Contexts of Special Pedagogy, 30. https://doi.org/10.14746/ikps.2020.30.11

9.0 ABOUT THE AUTHORS

The authors are fourth-year students pursuing Bachelor of Science degrees in Computer Engineering and Electronics Engineering at FAITH Colleges, located in Tanauan City, Batangas, Philippines. Throughout their academic journey, they have developed a strong foundation in engineering principles, research, and practical applications, preparing them for future careers in technology and innovation.

10.0 APPENDIX

Appendix A – Pre-Test and Post-Test Criteria

Category	Rating	Description
Speech Acceptability	0	Speech is acceptable
	1	Speech is mildly unacceptable
	2	Speech is moderately unacceptable
	3	Speech is very unacceptable
Speech Understandability	0	Typical or normal for age, understood except for rare instances
	1	Different from other children's speech, but not enough to affect understandability
	2	Comment provoked by difference in speech, but possible to understand at most lines
	3	Difficult to understand
	4	Impossible to understand
Hypernasality	0	None
	1	Minimal: slight increase in nasal resonance
	2	Mild: evident on high vowels
	3	Moderate: evident on all vowels
	4	Severe: evident on vowels and voiced consonants
Audible Nasal Air Emission	0	Absent
	1	Occasionally present
	2	Frequently present
Hyponasality	0	None
	1	Mid - partial denasalization of nasal consonants
	2	Marked - denasalization of nasal consonants & adjacent vowels
Voice	0	Normal
	1	Mild - Unusual/abnormal quality
	2	Moderate to severe - Unusual/abnormal quality

Appendix B – Performance Efficiency Criteria

Time behavior. Degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements.
Resource utilization. Degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements.
Capacity. Degree to which the maximum limits of a product or system parameter meet requirements.

Appendix C – Usability Criteria

Appropriateness recognizability. Degree to which users can recognize whether a product or system is appropriate for their needs.
Learnability. Degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use.
Operability. Degree to which a product or system has attributes that make it easy to operate and control.
User error protection. Degree to which a system protects users against making errors.
User interface aesthetics. Degree to which a user interface enables pleasing and satisfying interaction for the user.
Accessibility. Degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use.