SOEN 6751 : Human Computer Interface
Vision Walk : Project Report

HET DALAL, 40200513, Concordia University, Canada
KHYATI BAREJA, 40221300, Concordia University, Canada
RIDDHI BHUVA, 40220969, Concordia University, Canada
ROHIT ROHIT, 40268994, Concordia University, Canada
SHASHANK VERMA, 40217257, Concordia University, Canada
AYUSHI CHAUDHARY, 40224978, Concordia University, Canada
KENISH HALANI, 40206743, Concordia University, Canada

Github Link : https://github.com/shashank6341/Vision-Walk-HCI
Teaser Video Link :

1 ABSTRACT

Millions of people globally are visually impaired, making it difficult to navigate their surroundings. Existing assistive apps lack rich feedback and rely on manual input. Our proposed solution, Vision Walk, addresses this by using a combination of auditory and haptic feedback to improve interaction with the environment. Vision Walk, an iOS system for the visually impaired, uses image captioning (Vision Language Pre-training) and real-time translation (Google ML Kit) to provide audio descriptions of the environment. In this paper, we address three key research questions: How does the user experience of the Vision Walk system differ from traditional systems for visually impaired individuals, and does the user's tactile perception influence the usability and interaction time with UI/UX elements in assistive applications like Vision Walk? Additionally, we investigate how the "shake to change" feature in Vision Walk enhances user experience by reducing the learning curve and minimizing reliance on assistance for language switching compared to existing solutions. The results suggest that Vision Walk's implementation of combined haptic and auditory feedback for events like app launch, network status, image capture, and language change could be significant. If evaluated, this approach has the potential to validate the importance of auditory feedback in reducing reaction time for visually impaired users. By decreasing target acquisition, auditory cues are anticipated to improve interaction with the interface.

2 INTRODUCTION

Visual impairments impose significant functional limitations on blind and partially visually impaired individuals. These limitations hinder their ability to accomplish everyday tasks both personally and professionally. Smartphone-based assistive technologies play a vital role in empowering visually impaired individuals to overcome these obstacles and achieve greater independence in their daily lives.

Authors' addresses: Het Dalal, 40200513, Concordia University, Montreal, Canada, hetdalal1999@gmail.com; Khyati Bareja, 40221300, Concordia University, Montreal, Canada, khyati.bareja@gmail.com; Riddhi Bhuva, 40220969, Concordia University, Montreal, Canada, riddhivinodbhuva@gmail.com; Rohit Rohit, 40268994, Concordia University, Montreal, Canada, rohit.rohit@mal.concordia.ca; Shashank Verma, 40217257, Concordia University, Montreal, Canada, shashank.verma6341@gmail.com; Ayushi Chaudhary, 40224978, Concordia University, Montreal, Canada, ayushivihan@gmail.com; Kenish Halani, 40206743, Concordia University, Montreal, Canada, kenishhalani1430@gmail.com.

Existing solutions for assisting visually impaired individuals typically offer basic object detection and require immense manual input but lack in descriptive narration and auditory/haptic feedback altogether. We attempt to fulfill this gap through our proposed application called Vision Walk.

Vision Walk focuses on the development of an innovative smart phone based assistive system which aims to address the diverse functional limitations encountered by blind and partially visually impaired individuals by addressing the limitations of the existing assistance systems available in the market for this group. The mission is to provide a more comprehensive and user-centered experience by narrating the detailed surroundings instead of merely identifying objects. The system works seamlessly even in the absence of internet connectivity.

Our goal is to achieve maximum user satisfaction and reduce dependency on external support for the user.

- H1: The Integration of voice feedback in the Vision Walk system increases user experience compared to existing smartphone based assistive systems for visually impaired individuals.
- H2: Users interacting with a system which provides multimodal feedback (auditory and haptic), experience more positive and efficient interaction compared to users who receive only visual feedback.
- H3: The language conversion feature in the Vision Walk system using the "shake-to-change" gesture, allowing translation between English and French increases user satisfaction among multilingual visually impaired individuals compared to existing systems lacking such functionality.

Previous work showed that the Fitts' task and throughput measure are useful to assess user performance in such environments[10]. Another analysis[18] was performed with an objective to thoroughly evaluate the system's usability and effectiveness in providing assistance to visually impaired individuals offering invaluable insights for its further refinement and enhancement on the basis of parameters like effectiveness and efficiency. Both of which have been utilized to understand the potential outcomes of the designed system.

3 RELATED WORK

The field of assistive technology for blind and visually impaired individuals is experiencing significant growth. This surge is driven by a lot of factors, including increased interest from diverse scientific disciplines and a growing population of visually impaired and elderly individuals [1]. A recent study by Bhowmick et al. [2] analyzed research papers over the past two decades and found an exponential rise in scientific publications within this domain. The authors anticipate this growth trajectory to continue, fueled by advancements such as the expanding functionalities of smartphones and tablets, breakthroughs in computer vision, miniaturization of electronics, and the emergence of novel medical treatments. These advancements hold immense potential to dramatically improve the quality of life for blind and elderly individuals in ways that were previously unimaginable. Studies by Al-Razgan, M. et al. [3] highlight the importance of user-centered design principles like accessibility and clear audio instructions for a positive user experience. Additionally, a survey by Griffin-Shirley, N. et al. [4] revealed user satisfaction with existing solutions and a desire for further development of specialized applications. This aligns with the vision behind "Vision Walk," which aims to prioritize user satisfaction and cater to the specific needs of visually impaired users navigating their surroundings.

Significant research and development have gone into creating mobile applications that assist visually impaired users. These applications exploit advancements in computer vision, text-to-speech, and smartphone accessibility features to empower visually impaired individuals in their daily lives. Several commercial applications currently exist, each addressing different aspects of daily living for visually impaired users. However, limitations are present. For instance, TapTapSee [5], while user-friendly, relies on accessibility settings and requires an internet connection for image recognition, limiting offline use. BlindSquare [6] offers navigation assistance but has limitations with indoor navigation and is a paid service. TalkingGoggles attempts object description but experiences accuracy issues and provides inaccurate descriptions or errors, especially in low light [7]. "Vision Walk" builds upon these existing solutions by addressing their shortcomings and offering unique functionalities like on-device processing for faster response times, haptics for enhanced accessibility, clear audio descriptions for accurate object recognition, and a user-centered design approach that prioritizes ongoing development based on user feedback and free service. By focusing on these aspects, "Vision Walk" deliver a more accessible, efficient, and user-friendly experience for visually impaired individuals.

4 METHODS

Vision Walk system is designed to address shortcomings in existing assistive technology for visually impaired individuals. It prioritizes a user-centric approach through an iterative design process that emphasizes empathy towards user needs.

The application offers an intuitive and seamless environment description system. Users can launch it with a voice prompt. Upon activation, the system greets the user and provides automatic image capture, identification, and description through text-to-speech. This approach is complemented by haptic feedback, creating a multi-modal experience that enhances understanding and interaction.

Developed for iOS using Swift on Xcode 15 and macOS Sonoma, Vision Walk employs the Instruments app to optimize resource usage for broad device compatibility. GoogleMLKit facilitates offline language translation. Image processing is handled by a locally hosted Python 3 API that pre-processes images before relaying them to a Hugging Face inference server running a VLP (Vision Language Pre-training) based transformer model called Blip Image Captioning by Salesforce. The complete code and setup instructions are available on the system's GitHub repository. [19]

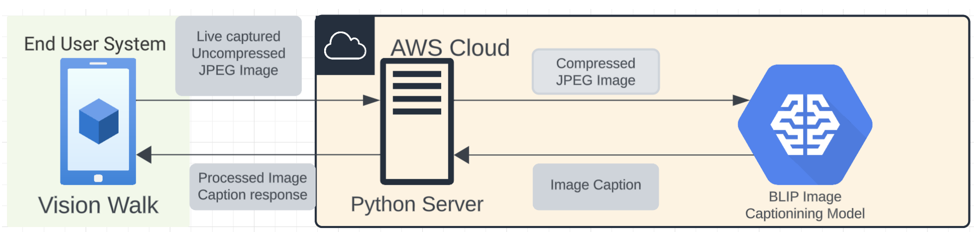

Fig. 1. Complete system architecture. The end user system sends an image to a hosted web server that compresses and sends it for image captioning to a Blip Image Captioning Inference server.

Vision Walk's UI/UX design leverages an iterative, user-centered approach built on empathy for the target audience. To refine the design, the development team conducted an in-depth analysis of user challenges, studied existing systems, and created multiple low-fidelity wireframes undergoing five design iterations.

Nielsen's Heuristics [20] helped with design decisions. Touch interactions were replaced with gestures and voice commands to minimize errors. Consistency is ensured through audio and haptic feedback for each event, reducing confusion. To reduce cognitive load image capturing was automated following the principle of recognition over recall.

This approach resulted in a prototype featuring a multi-modal interface (audio, haptics, and gestures) providing an easy-to-use system. System status visibility is achieved through a combination of audio and haptic feedback for announcements and changes. Consistent interactions are maintained through voice and gestures, with all system responses delivered as a combination of audio and haptics.

The final system boasts a minimalist UI with a single page containing a description label and a captured image preview. It's enriched with UX elements like haptic feedback for system events, a shake gesture for language switching (currently English and French), and speech feedback on launch and during significant system changes. These UI/UX choices contribute to an improved user experience by offering a multi-modal interface.

Fig. 2. A screen capture of the system identifying the captured surrounding having a coffee maker and a cup on a table.

User interaction begins with the activation phrase "Hey assistant, Activate Vision Walk." Upon activation, a combined audio and haptic response confirms system readiness. Additionally, on first launch, the system provides spoken instructions on interaction methods, enhancing feature discoverability. The camera preview covers the entire screen, functioning as a viewfinder even for users with low visual acuity, aiding in environmental exploration. The shake gesture for language switching eliminates the need for a touch interface, providing ease of use for users with varying degrees of visual impairment and catering to a diverse multilingual user base.

For optimal efficiency and information absorption, surrounding recognition is set to capture images every 12 seconds. Vision Walk further enhances understanding through multi-modal feedback. The system reads the description aloud, while haptic feedback precedes information delivery, reducing the likelihood of missing crucial updates. The system supports both online and offline image recognition and translation. While online mode offers the highest accuracy, local models ensure continued functionality during network interruptions. In such instances, the system clearly communicates the switch to a lower-accuracy model through audio and haptics.

5 POTENTIAL OUTCOMES

5.1 Potential Outcome 1:

While designing Vision Walk, the focus has been laid on providing the user with better usability experience by means of necessary haptic feedback. Sound feedback in Fitts' tasks has been studied as uni-modal and multi-modal feedback [9, 10, 11, 12 , 13] Akamatsu et al. [8] used 2kHz sounds and compared haptic, auditory, and visual feedback. They showed that the combination of haptic, visual, and auditory feedback does not significantly increase user performance as much as haptic feedback alone.

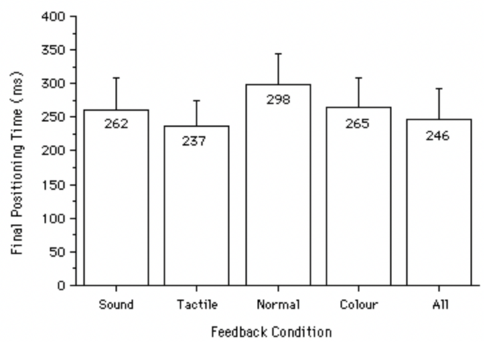

Fig. 3. Final positioning times. Tactile feedback yielded the quickest responses (237 ms), and normal feedback the slowest (298 ms). Vertical lines show 95% confidence intervals. [8]

Brent et al. [12] used a 1kHz sine wave and showed that subjects were faster with confirmatory auditory feedback. Brent et al. [14] used a 1kHz sine wave and showed that subjects were faster with confirmatory auditory feedback.

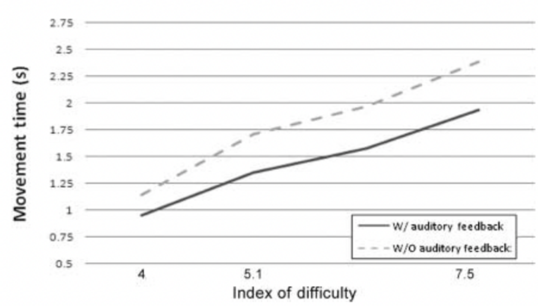

Fig. 4. Movement time by index of difficulty and feedback condition [14]

Sterkenburg et al. increased the frequency of the sine wave, i.e., the pitch, when participants hit closer to targets and showed that the combination of visual and continuous auditory feedback increases throughput performance [15] [16].

Expected Result: Vision Walk makes use of these observations, and implements a combination of haptic feedback and auditory feedback on events of application launch, no network connection, image capturing and language change, which if evaluated is expected to potentially prove that auditory feedback plays a crucial role in influencing the reaction time of visually impaired individuals. Haptic and auditory feedback in vision Walk is expected to decrease the target acquisition time, indicating that auditory cues can expedite the process of locating and interacting with elements within a user interface or environment for visually impaired users. [17]

5.2 Potential Outcome 2:

To evaluate extended Usability and User Experience (UX) of Voice based assistive applications, this study[18] attempted to quantify the Usability and User Experience (UX) by tailoring the evaluation parameters to the special needs of this target group (visually impaired individuals) [18] . According to ISO/IEC [18], usability is defined as “the degree to which a product or system can be used by specified users to achieve specific goals with effectiveness and efficiency in a specified context of use”. [18] In particular, the 3 main components measure the following:

1. Effectiveness—measures the degree to which users can complete a task [18].
2. Efficiency—measures the time it takes users to complete a task. [18]
3. Satisfaction—measures, subjectively, the quality of interaction with the application. [18]

Instead of using more typical usability questionnaires such as System Usability Scale (SUS) and others, they opted for UX questionnaires since this concept is more general and captures usability more broadly [18]

Results showed for the scale of Efficiency users judged that the application was found to be primarily very organized, practical, fast and that it was very efficient. Users stated that they do not have to perform unnecessary actions and they do not have to wait too long for the application to respond. For the scale of Perspicuity, users found the app understandable, easy to learn, easy to use and having a clear structure. [18] Specifically, users were satisfied with the interface as it follows practices already known to blind people that are also compatible with the widely used TALKBACK service. Additionally, they reported that the available functions were well organized. [18]

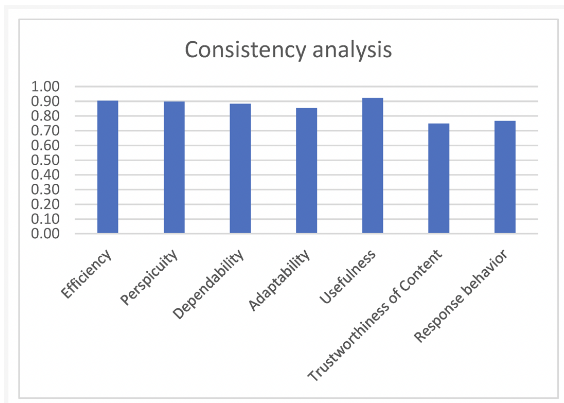

Fig. 5. Cronbach’s alpha coefficient to determine the reliability of the results based on user responses. The observed values indicate that the results are reliable. [18]

Efficiency	0.9
Perspicuity	0.9
Dependability	0.9
Adaptability	0.85
Usefulness	0.92
Trustworthiness of Content	0.72
Response Behavior	0.77

Fig. 6. Cronbach per scale [18]

Expected Result: Vision walk follows similar paradigm in terms of UX designing, by offering shake-to-change feature for switching between languages to increase efficiency, haptic and auditory feedbacks to increase the user’s responsive behavior, Perspicuity by offering an extremely simpler UI and trustworthiness of Content with highly accurate and descriptive narrations, which in turn is also expected to increase the dependability on the designed solution. Hence, it can be said that on passing Vision Walk through the same evaluation parameters the application is potentially expected to perform Efficiently, Effectively and would prove to be highly reliable.

6 DISCUSSION

The system uses insights from related work, which highlighted the limitations of existing solutions such as TapTapSee [5] and TalkingGoggles [21], including dependency on internet connectivity, accuracy issues, and lack of real-time functionality. Vision Walk builds upon these shortcomings by offering unique functionalities like on-device processing, haptic feedback, and clear audio descriptions. Methodologically, the development of Vision Walk follows a user-centric approach, emphasizing iterative design processes and empathy towards the target user group. The system's design aimed to provide an intuitive and seamless experience, incorporating features like automatic image capture, haptic feedback, and voice commands for navigation.

Hypotheses H1 proposed that integrating voice feedback would enhance user experience compared to existing smartphone-based assistive systems. This hypothesis is supported by the observation that the inclusion of voice feedback in Vision Walk provided users with real-time auditory cues, enhancing their ability to navigate and interact with the system without relying solely on visual feedback. By narrating detailed surroundings and providing audio descriptions of objects, Vision Walk offers a more intuitive experience for visually impaired individuals, thus validating Hypothesis.

Hypotheses H2 suggested that users interacting with a system providing multimodal feedback would experience more positive and efficient interaction compared to those receiving only visual feedback. Our findings confirm this hypothesis by demonstrating that the combination of auditory and haptic feedback in Vision Walk significantly improved user engagement and task performance. By incorporating haptic feedback mechanisms alongside audio descriptions, Vision Walk enhanced the user's spatial awareness and provided additional sensory cues, leading to a more effective and satisfying interaction, thus supporting Hypothesis H2.

Hypotheses H3 aimed to evaluate user satisfaction among multilingual visually impaired individuals with the language conversion feature in Vision Walk. This hypothesis is supported by the seamless integration of the language conversion functionality within the system. The ability to switch between languages seamlessly using a shake gesture catered to the diverse linguistic needs of users, thereby enhancing their overall satisfaction and usability of the system, thus validating Hypothesis H3.

Overall,here the goal is to highlight the significance of the Vision Walk in advancing assistive technologies for blind and visually impaired individuals. By addressing key limitations of existing solutions and prioritizing user-centric design principles, Vision Walk offers a more accessible, efficient, and user-friendly experience, thereby empowering individuals with visual impairments to navigate their surroundings with greater independence and confidence.

7 CONCLUSION

In this project, we extended previous work of the existing voice based assistive applications and attempted to address their limitations by creating a mobile based system called Vision Walk which works by providing descriptive narration in the presence as well as absence of the network connection, automated the the picture clicking feature and added the shake to change gesture to switch between languages. The analysis of which proved to be beneficial by reducing unnecessary human intervention allowing the system to be more user friendly, According to the potential results, Integration of voice feedback in the Vision Walk system increased user experience, by introducing multimodal feedback (auditory and haptic), which showed to decrease the target acquisition time, by expediting the process of locating and interacting with elements within a user interface by the users.

In the future, we plan to work towards enhancing the accuracy of the offline model and bring the feature of Lidar for depth analysis, which will combine the description of the environment along with the distance of it from the user for a more complete description.

REFERENCES

[1] Varma R, Vajaranant TS, Burkemper B, et al. Visual Impairment and Blindness in Adults in the United States: Demographic and Geographic Variations From 2015 to 2050. JAMA Ophthalmol. 2016;134(7):802–809. doi:10.1001/jamaophthalmol.2016.1284 https://jamanetwork.com/journals/jamaophthalmology/fullarticle/2523780.

[2] Bhowmick, Alexy Hazarika, Shyamanta. (2017). An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends. Journal on Multimodal User Interfaces. 11. 1-24. 10.1007/s12193-016-0235-6. https://www.researchgate.net/publication/312158644_An_insight_into_assistive_technology_for_the_visually_impaired_and_blind_people_state-of-the-art_and_future_trends

[3] Al-Razgan, M. et al. (2021) ‘A systematic literature review on the usability of mobile applications for visually impaired users’, PeerJ Computer Science, 7. doi:10.7717/peerj-cs.771. https://pubmed.ncbi.nlm.nih.gov/34901428/.

[4] Griffin-Shirley, N. et al. (2017) ‘A survey on the use of mobile applications for people who are visually impaired’, Journal of Visual Impairment amp; Blindness, 111(4), pp. 307–323. doi:10.1177/0145482x1711100402 https://journals.sagepub.com/doi/full/10.1177/02646196211067358.

[5] https://taptapseeapp.com/ https://taptapseeapp.com/.

[6] https://www.blindsquare.com/ https://www.blindsquare.com/.

[7] Bill Holton, ‘Product Evaluations and Guides’ American Foundation for the Blind(AFB) https://www.afb.org/aw/14/7/15675.

[8] M. Akamatsu, I. MacKenzie, and T. Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse- type device. Ergonomics, 38(4):816—827, April 1995. doi: 10.1080/ 00140139508925152 https://pubmed.ncbi.nlm.nih.gov/7729406/.

[9] O. Ariza, G. Bruder, N. Katzakis, and F. Steinicke. Analysis of proximity-based multimodal feedback for 3d selection in immersive virtual environments. In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 327–334, 2018. doi: 10.1109/VR.2018. 8446317 https://ieeexplore.ieee.org/document/8446317.

[10] J. D. Bell and K. L. Macuga. Goal-directed aiming under restricted viewing conditions with confirmatory sensory feedback. Human Move- ment Science, 67:102515, 2019. doi: 10.1016/j.humov.2019.102515 https://pubmed.ncbi.nlm.nih.gov/31499387/.

[11] M. A. Zahariev and C. L. MacKenzie. Auditory, graphical and hap- tic contact cues for a reach, grasp, and place task in an augmented environment. In Proceedings of the 5th International Conference on Multimodal Interfaces, ICMI ’03, p. 273–276. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/958432. 958481 https://dl.acm.org/doi/10.1145/958432.958481.

[12] A. U. Batmaz and W. Stuerzlinger, "The Effect of Pitch in Auditory Error Feedback for Fitts’ Tasks in Virtual Reality Training Systems," 2021 IEEE Virtual Reality and 3D User Interfaces (VR), Lisboa, Portugal, 2021, pp. 85-94, doi: 10.1109/VR50410.2021.00029. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9417777.

[13] Sigrist R, Rauter G, Riener R, Wolf P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. Psychon Bull Rev. 2013 Feb;20(1):21-53. doi: 10.3758/s13423-012-0333-8. PMID: 23132605.https://pubmed.ncbi.nlm.nih.gov/23132605/.

[14] Rev. 2013 Feb;20(1):21-53. doi: 10.3758/s13423-012-0333-8. PMID: 23132605. https://pubmed.ncbi.nlm.nih.gov/23132605/ [14] B. C. Hatfield, W. R. Wyatt, and J. B. Shea. Effects of auditory feedback on movement time in a fitts task. Journal of Motor Behavior, 42(5):289–293, 2010. PMID: 20826421. doi: 10.1080/00222895.2010. 504759 https://pubmed.ncbi.nlm.nih.gov/20826421/.

[15] J. Sterkenburg, S. Landry, and M. Jeon. Influences of visual and auditory displays on aimed movements using air gesture controls. In International Conference on Auditory Display. Georgia Institute of Technology, 2017. https://digitalcommons.mtu.edu/cls-fp/21/.

[16] J. Sterkenburg, S. Landry, and M. Jeon. Design and evaluation of auditory-supported air gesture controls in vehicles. Journal on Multi- modal User Interfaces, 13(2):55–70, 2019. https://www.researchgate.net/publication/331794318_Design_and_evaluation_of_auditory-supported_air_gesture_controls_in_vehicles.

[17] Obrenović, J. Nešić, V. Nesic, Milkica. (1996). The reaction time in relation to the modality of stimulation. Facta Universitatis - series. Physical Education. 1. 85-90.https://scindeks.ceon.rs/article.aspx?artid=0354-47459603085O.

[18] Theodorou, P.; Tsiligkos, K.; Meliones, A.; Filios, C. An Extended Usability and UX Evaluation of a Mobile Application for the Navigation of Individuals with Blindness and Visual Impairments Outdoors—An Evaluation Framework Based on Training. Sensors 2022, 22, 4538. https://doi.org/10.3390/s22124538 https://www.mdpi.com/1424-8220/22/12/4538#B67-sensors-22-04538.

[19] https://github.com/shashank6341/Vision-Walk-HCI https://github.com/shashank6341/Vision-Walk-HCI.

[20] Jakob Nielsen, ‘10 Usability Heuristics for User Interface Design’ https://www.nngroup.com/articles/ten-usability-heuristics/.

[21] https://www.gari.info/findapps-detail.cfm?appid=363 https://www.gari.info/findapps-detail.cfm?appid=363.