Case Study: JOY Develops the First Musical-Visual Instrument with the Intel® Perceptual Computing SDK

By Karen Marcus

TheBestSync, a development company based in China, is a creative and energetic team with years of production experience. The company’s focus is on providing comprehensive technological solutions and execution integrated in an artistic package. Two technologies they have used to achieve these ends are perceptual computing and augmented reality (AR) technology, which is the integration of digital information with live video or the user’s environment in real time. AR also uses location-based system information to enhance users’ expression and sense of identification.

The team has created a few applications—mostly shooting and racing games—using Kinect® technology. For example, one of the games involves two people racing, using their hands to try and wave the fastest.

TheBestSync’s participation in Phase I of the Intel® Perceptual Computing Challenge offered the opportunity to create a musical-visual instrument based on perceptual computing. The application, called JOY, is intuitive enough for anyone (even children) to use, including those without musical training. The team was excited to take their perceptual computing and AR experience and knowledge to the next level.

A Musical Innovation: JOY

JOY is the first perceptual music-visual instrument. Performers can use it to display different sound elements and visual effects by altering the gestures, distance, depth, and altitude of their hands in front of the Creative® Interactive Gesture Camera (the Camera). The result is a simultaneous audio and visual experience operated through simple performer control.

JOY was conceived by Alpha Lam, CEO of TheBestSync. With experience as a sound engineer and musician, Lam wanted to develop an instrument that allows users to perform music without the tough learning curve. In researching perceptual computing technology, he thought of designing an instrument that enables users to just move their fingers, without touching any physical object, to play. He called together music and programming specialists within TheBestSync to work on this project. Getting started was a challenge, but after several rounds of testing, the team became convinced of the advantages of the Intel® Perceptual Computing Software Development Kit (SDK) as well as the new method of playing music based on actions anyone could make.

In his vision for JOY, Lam sees users playing music at home, sharing their creativity with friends, or showcasing it at parties. In particular, says Lam, “Professional DJs or musicians can feel free to express their unlimited musical creativity on stage.” (See Figure 1.)
Development Using the Intel Perceptual Computing SDK

As perceptual computing technology advances, gesture, facial, and voice recognition will fundamentally change how users interact with computers. At the time of the challenge, the Intel Perceptual Computing SDK was still in beta, and the plan was to take participants’ feedback to improve future releases of the SDK.

When TheBestSync developers heard about Intel’s perceptual computing innovations, they saw a match with other software they were developing. The team conducted an in-depth study of the beta version of the Intel Perceptual Computing SDK, and decided to join Phase I of the Intel Perceptual Computing Challenge. Lam says, “We created JOY based on the advantages of perceptual computing and hope more people can get to know our design to promote perceptual computing instrument development.”

JOY was designed specifically for the Challenge (see Figure 2). Lam says, “We considered the status of the Intel Perceptual Computing SDK and fully used the gesture control function. We hadn’t used the same range of gestures in previous apps; those developed for JOY were new for us. We wanted to let users get to know the advantages of perceptual computing through our application.”

Figure 1. JOY in use
Figure 2. A menu screen in JOY

Deciding Which Elements to Include

The team tried various input modes during the development process, including face, gesture, and voice modules. Lam explains, “When users tilted their head or turned it to the left or right, the face module sometimes failed to recognize the face. In this case, the users would need to turn their head back to the front to reactivate the recognition. I believe Intel is trying hard to improve this issue in the SDK. We used a program algorithm to improve this but still weren’t able to fully solve the problem.”

Lam adds, “For the voice module, the recognition was a bit slow. The new version of the SDK improves the voice recognition a lot, including increased language options and recognition capability.”

In the version of the application submitted for the Challenge, gesture recognition was the only input mode used. Lam notes, “We tested the application and found that the gesture control was the most stable part, the part that staged the best, and the easiest part for users to control. Therefore, we designed the application control mode based on hand manipulation and music manipulation. We modified the creative direction based on making the experience of using the application as user-friendly as possible.”
The initial idea for gesture-controlled functions came from a combination of Lam’s understanding of music performance and his knowledge of perceptual computing. Application improvement came from user testing. Lam notes, “In our early efforts with gesture recognition, there were errors in left/right hand recognition. For example, the default setting was for the first hand recognized to be the left hand. So, when players raised their right hand first to start the game, there was a recognition error, and all following actions got swapped between the right and left hands. To resolve this, we added criteria to assist with recognition. For example, we programmed the application to compare the elbow identification point versus the palm identification point; if the x-axis coordinates of the elbow identification point were bigger than the palm point, it was seen as the right hand. We realize this solution may still need to be improved.”

In addition, there were miscalculations as to how many fingers were being captured. The team filtered the finger quantity calculation to ensure stable recognition.

To determine which gestures worked best, the team did several rounds of testing to ensure that each one was stable and able to control the application continuously and individually. “For each gesture,” says Lam, “we tried to make it as intuitive and easy to associate as possible. For example, changing the distance between right and left hands horizontally activates reverb, while changing it vertically activates an echo effect.” The team found that five fingers all open worked best for recognition; waving or circling hands was also stable.

The interface used to show how many fingers are being captured was inspired, says Lam, by stage lights: “We tried our best not to destroy the overall visual aesthetic by showing the fingers while providing clear enough hints to players.”

**Design Challenges**

The biggest challenge in the development process was making the finger controls more user-friendly. The team implemented several adjustments:

- To ensure the application accurately detected the numbers of validated fingers, the team leveraged a screening technique to filter out the unstable part and provide precise finger detection. A change to sampling frequency was not needed, but, says Lam, “If we screened three frames with the same result, it was confirmed as an effective recognition, and the application filtered out the invalid data.”
- To control the mix of gestures and the order in which they occur, the team connected musical tracks to the number of fingers recognized. The fingers captured activate corresponding musical quantities and sequences.
- To maintain effective program fluency when changing gesture movements, the team filtered data to reduce recognition errors. They also added a preliminary judgment, which enables the application to judge which function the user is controlling.
**Testing**

To test this experience, the team invited two different groups of people to try the application: users with no musical background and professional musicians. They wanted to ensure that the application was user-friendly enough for those with a limited music background.

In a simple introduction that explained the way the application works, the team told testers: “Each of the 10 fingers represents a track of music, so, 10 tracks are possible to use for creating different sounds. Users can remix them through changes to finger combinations.” Following this introduction, both groups of users could easily manipulate the application.

The musician testers expected more functions and more ways of manipulating the application. Conversely, the typical response from those without musical knowledge was that it was a cool application and a brand new experience.

**Future Plans**

Though the team developed the application for the Challenge, they continue to improve it. Lam explains: “In the new version, we added a face landmark application programming interface (API) to use different input modes. We plan to add touch screen and keyboard input to enable users to switch between different devices and to provide a smoother, easier manipulation experience.” Lam adds, “We will enable the application to switch from gesture to touch screen, and we will switch the status.” The application will also switch to touch mode when users intentionally change from gesture input to the touch screen, but if no follow-up action is taken after a switch is detected, the application will remain in the same mode (see Figure 3).

![Figure 3. JOY with facial recognition](image)

In addition, the team has added music factor recording—an import function—music play list editing, and music sharing.
The team’s goals for JOY include popularizing Camera functionality and commercializing the application.

Development Experience

The team is pleased with their success in applying the perceptual music and visual playing concept to JOY. Lam believes this is the first time these components have all been brought together in one application. Another achievement was redefining the perceptual instrument to synchronize audio and visual experiences.

They learned some valuable lessons through the development experience. As a result of the development, says Lam, “TheBestSync now has an in-depth understanding of the different perceptual computing APIs as well as the lessons we learned during the testing process. Now we can design and create applications based on what we know about the advantages and disadvantages of perceptual computing.” He adds, “We have developed some AR applications for user interaction, such as one that can ‘see through’ the laptop (like a computer X-ray) to create a fun environment for users to interact with and better showcase products.”

The team urges other developers to dig into the details to understand the platform and conduct tests during the idea-generation stage to showcase the advantages of perceptual computing and avoid the disadvantages. Lam notes, “The Intel Perceptual Computing SDK is still under development, and we can foresee additional functional improvements in the future. Keeping an eye on the latest updates is crucial to development.”

Lam reflects that the emergence of perceptual computing offers more possibilities for music performance and creation. He says, “It’s a brand new experience in music performance. Anyone who likes music, even without instrument or music knowledge, can flexibly, simply, and intuitively manipulate sound using the perceptual interface, then create and perform various styles of music. It is a true music-creation tool that has no boundaries.”

Further, says Lam, “We see perceptual computing as a more intelligent human-computer interaction. It allows the computer to ‘sense’ users, while users have a more intuitive and natural way to manipulate the computer. It’s similar to going from using a mouse to using a touch screen; this is another innovation in input methods. We believe perceptual computing will redefine the relationship between human and computer, enable more input mode possibilities, and bring the user experience into a brand new world.”

Other Development Considerations

The team used the Unity* 3D development engine to quickly integrate three-dimensional (3D) models and animation into the application. Lam notes, “The Unity Asset Store also provides sufficient plug-ins to make the development more efficient.”
To program JOY, the team used C# and C++. They chose C# because it is compatible with Unity 3D. Lam says, “C# can quickly and conveniently adapt to Unity development. C# provides data encapsulation to raw data, which is convenient for developers and shortened our development duration.”

C++ was selected because the Unity interface included in the Intel Perceptual Computing SDK was insufficient to fulfill development needs. For example, says Lam, “There was no prompt function when users’ hands exceeded the detectable range. We found that age, gender, and facial expression functions were also lacking.” So, the team used C++ to extend the Unity API beyond the Intel Perceptual Computing SDK and complete the development. They found that it provided convenient function extension.

Other tools included:
- Autodesk Maya* 3D animation, which has powerful design functions that are easily adaptable to Unity
- Avid Pro Tools|HD, which has powerful sound recording and editing functions that provide a higher sound quality
- Apple Logic Pro, which enables flexible music design to establish sufficient sound resources
- NGUI Unity plug-in, which efficiently brought user interface (UI) artistic design to the program

Company

Previous TheBestSync application development covered numerous apps, including AR apps, games, UI designs, and web sites. As for the future, TheBestSync will continue to advance the development of perceptual computing by increasing its labor investment, inviting third-party investors, and using other resources to support long-term development. Lam says, “With our strong production background and advanced technology development, we aim to provide a perfect user experience by integrating perceptual technology into innovative applications.”

The company is also participating in Phase 2 of the Intel Perceptual Computing Challenge.

To learn more about TheBestSync, go to www.thebestsync.com.

About the Author

Karen Marcus, M.A., is an award-winning technology marketing writer who has 16 years of experience. She has developed case studies, brochures, white papers, data sheets, solution briefs, articles, web site copy, video scripts, and other documents for such companies as Intel, IBM, Samsung, HP, Amazon Web Services, Amazon Webstore, Microsoft, and EMC. Karen is familiar with a variety of current technologies, including cloud computing, IT outsourcing, enterprise computing, operating systems, application development, digital signage, and personal computing.
Notices
INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL PRODUCTS. NO LICENSE,
EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS
GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR
SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR
IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR
WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR
INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.

UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR
INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A
SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the
absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future
definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.
The information here is subject to change without notice. Do not finalize a design with this information.

The products described in this document may contain design defects or errors known as errata which may cause the product
to deviate from published specifications. Current characterized errata are available on request.

Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product
order.

Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be
obtained by calling 1-800-548-4725, or go to: http://www.intel.com/design/literature.htm
Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.
Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components,
software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult
other information and performance tests to assist you in fully evaluating your contemplated purchases, including the
performance of that product when combined with other products.
Any software source code reprinted in this document is furnished under a software license and may only be used or copied in
accordance with the terms of that license.
Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries.
Copyright © 2013 Intel Corporation. All rights reserved.
*Other names and brands may be claimed as the property of others.