



# MEET LIGHTSPEED STUDIOS AT GDC2023

March 20-24, 2023 | San Francisco, CA



# Investigating the Utilization of Gvoice AI Codec and Haptics Technology in Mobile Gaming to Enhance Player Experience

**Chao Weng**

Principal Researcher and Team Lead,  
LIGHTSPEED STUDIOS

**Shili Xu**

Game Technical Operation Director,  
LIGHTSPEED STUDIOS

March 20-24, 2023 | San Francisco, CA

# Outline

**PART.1**

**Gvoice AI Codec**

**PART.2**

**Haptic Feedback: How It Brings  
Inclusivity and Accessibility**



**PART.1**

# Gvoice AI Codec





# Gvoice AI Codec: Motivations

## Motivations:

- In-game **voice chat plays a crucial role** in the development of multiplayer gaming.
- **A versatile and robust speech codec** is essential for providing stable and high-quality in-game voice communication on various gaming devices.

## Challenges:

- A traditional speech codec typically requires **a high bit rate** to provide high-quality speech transmission.
- A **prohibitively high bandwidth cost** is incurred to ensure quality and accommodate a large number of gamers.
- Opus, the most commonly used speech codec (existing solution), **degrades dramatically** in poor network connections

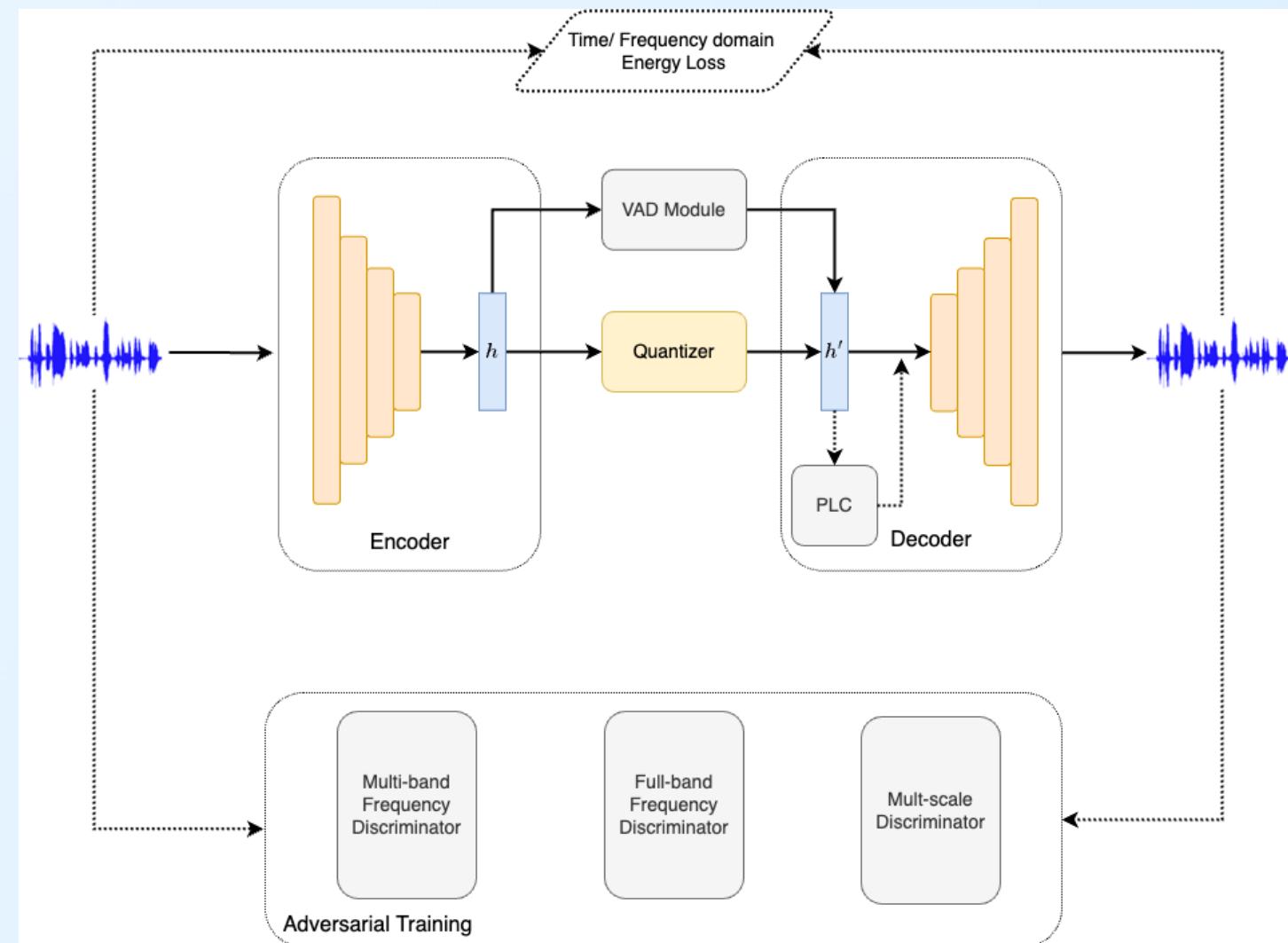
# Gvoice AI Codec: Distinctive Properties

- **Objectives :** high speech quality, high compression rate, low latency and complexity, good loss robustness
- **Gvoice AI Codec:** an end-to-end neural speech codec tailored for in-game chat on mobile devices
- **Distinctive Properties:**
  - lightweight footprint and low computational cost enabling deployment on various mobile edge devices
  - remarkable speech quality preservation and compression rate
  - neural packet loss concealment leading to good loss robustness without extra bit overhead
- **Gvoice AI Codec** has been fully adopted by some of the most influential mobile games, including **PEACEKEEPER ELITE( Game for Peace ), etc.**



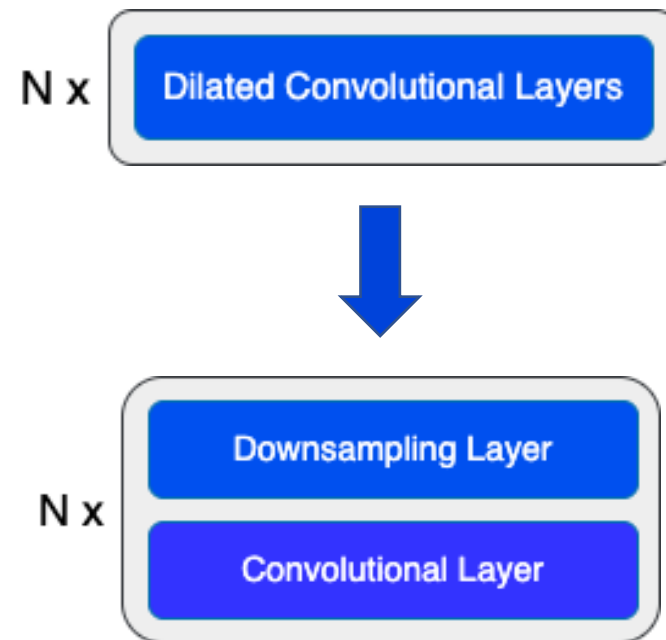
# Gvoice AI Codec: System Architecture

- **Gvoice AI Codec** replaces traditional Opus with neural network based compression modules.
- **Encoder:** encodes and downsamples the speech signals to generate embeddings (real-valued vectors)
- **Quantizer:** discretizes the embeddings
- **Decoder:** upsamples and decodes the speech signals
- All neural network modules are optimized using a **reconstruction and GAN-based loss** function in an end-to-end fashion



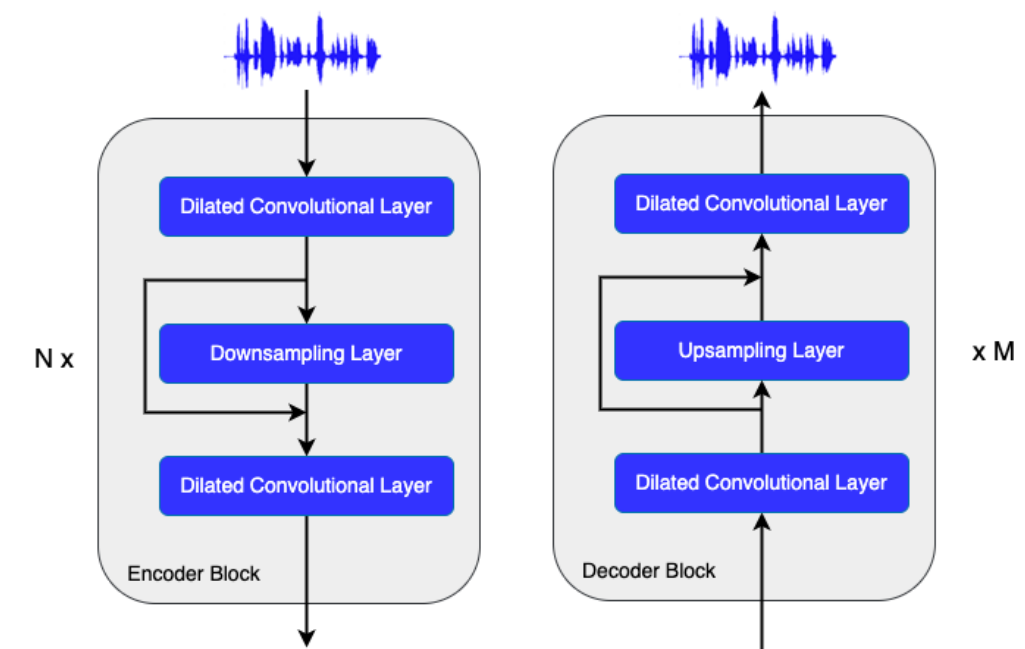


# Gvoice AI Codec: Major Components – Neural Network Model Design



## Stack Downsampling and Convolutional Cascaded Layers

- effectively reduce the computational load
- convolutional layers with the same amount of computation can obtain a wider receptive field



## Residual Connections

- preserve local context
- effective fusion of local and global contextual information
- improve the sound quality
- lower the computational complexity



# Gvoice AI Codec: Major Components – Model Training Strategies

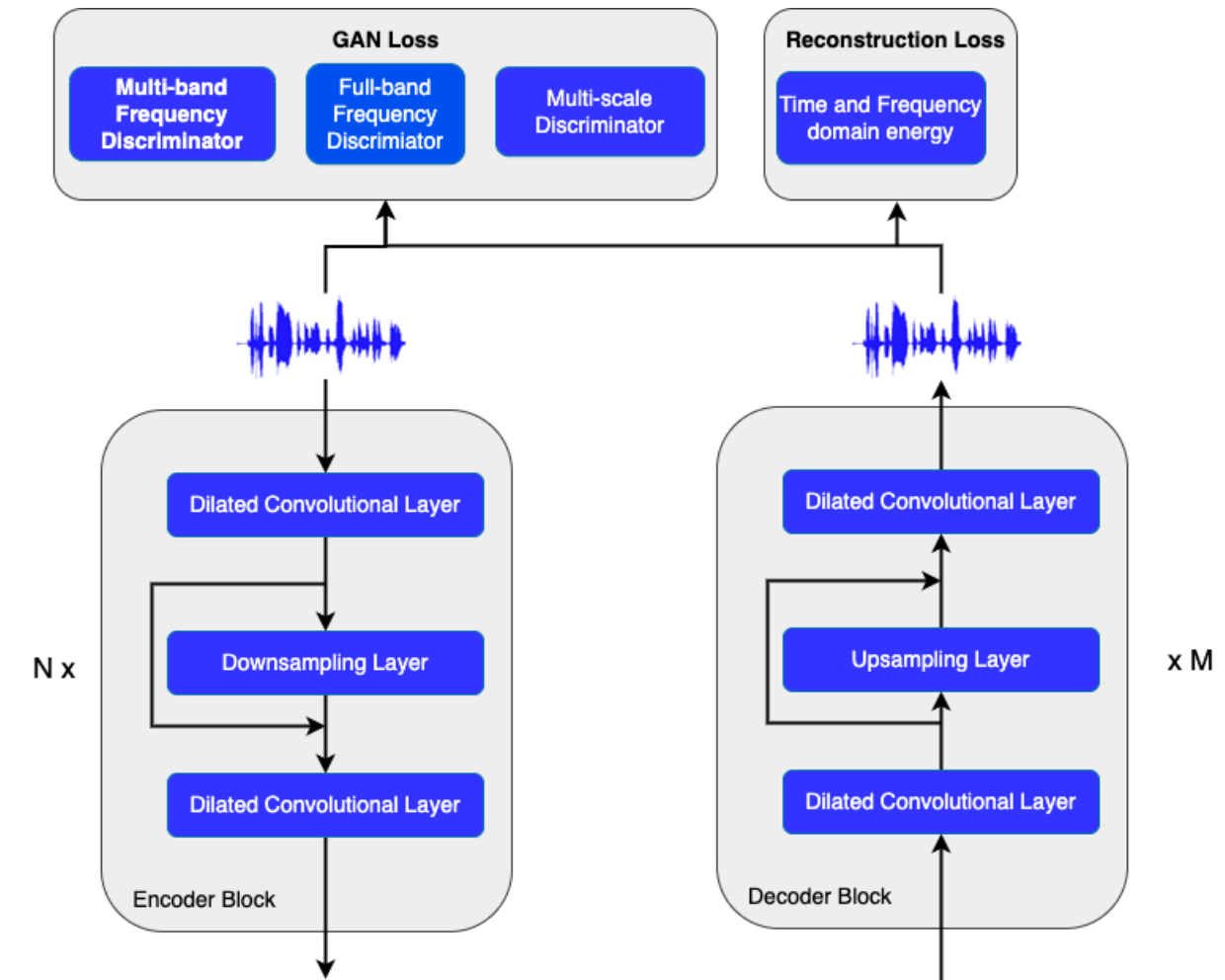
## Reconstruction Loss :

## Adversarial Loss:

- multi-scale and multi-resolution adversarial loss
- introduces the sub-band discriminator of the frequency band (few if any)
- improves the restored voice quality without increasing the overall network parameters and calculations

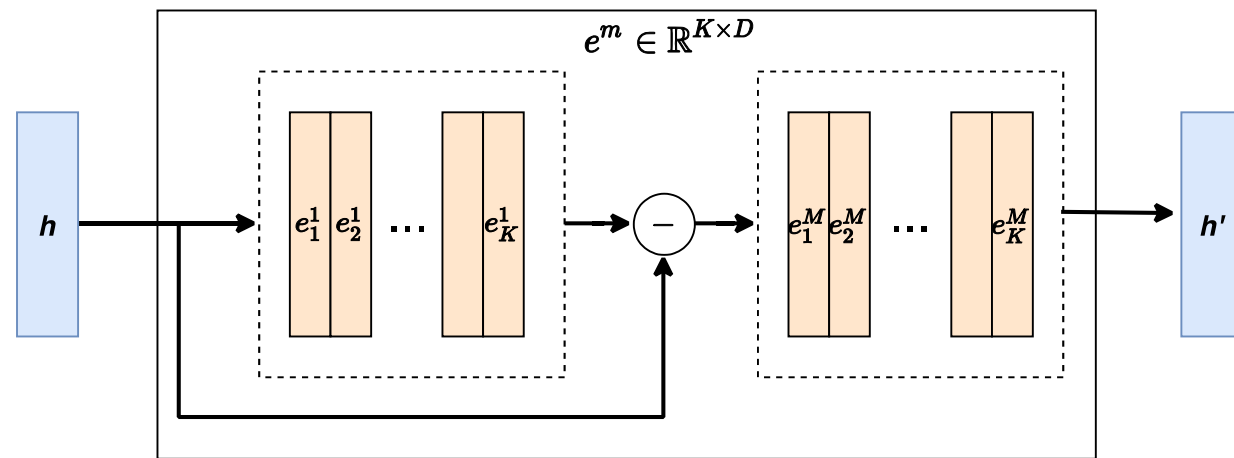
$$\mathcal{L}_{adv} = \sum_{k=1}^K (D_k(G(x)) - 1)^2 + \sum_k \sum_i \frac{1}{N_i} \|D_k^i(x) - D_k^i(G(x))\|_1$$

$$\mathcal{L}_{energy} = \|f_{time}(x) - f_{time}(\hat{x})\|_1 + \|g_{frequency}(x) - g_{frequency}(\hat{x})\|_1$$



# Gvoice AI Codec: Major Components – Codebook-free Quantizer

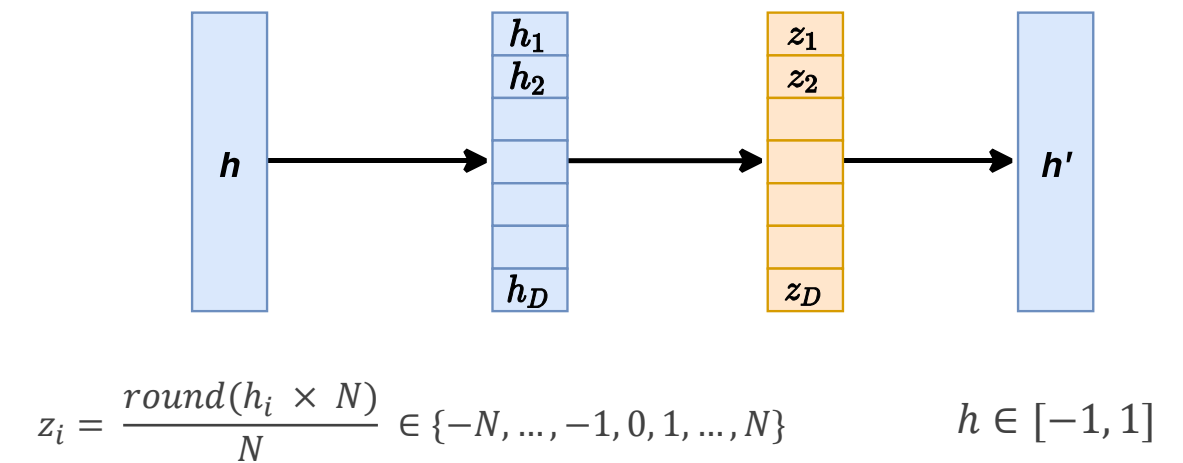
## Residual Vector Quantization



- The quality heavily depends on the complexity (e.g., size and number) of the codebooks
- Additional memory is required.
- A typical configuration: 8 codebooks, each with a size of 1024 and a dimension of 128  $\rightarrow$  1M floats

## Scalar Quantization with residual entropy coding

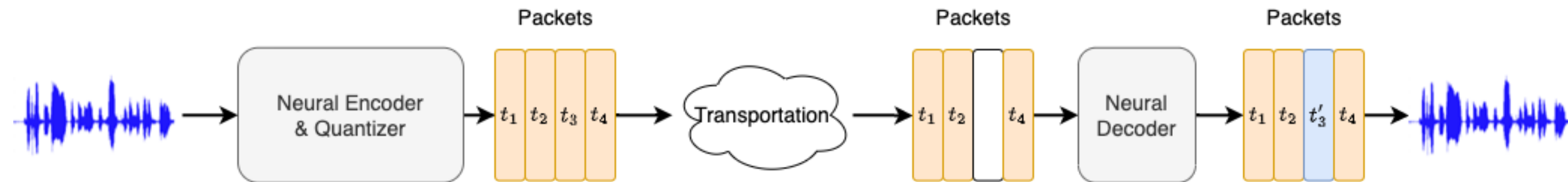
(this work)



- Scalar Quantization:
- No additional memory overhead for codebooks
- Entropy coding (e.g., Huffman, Range) after quantization
- Slight changes over consecutive speech packets, esp. for silence or stationary noise segments
- Further reduce bits via entropy coding on residual signal (differences between consecutive packets)
- Additional benefits: residual information also helps determine voice/silence



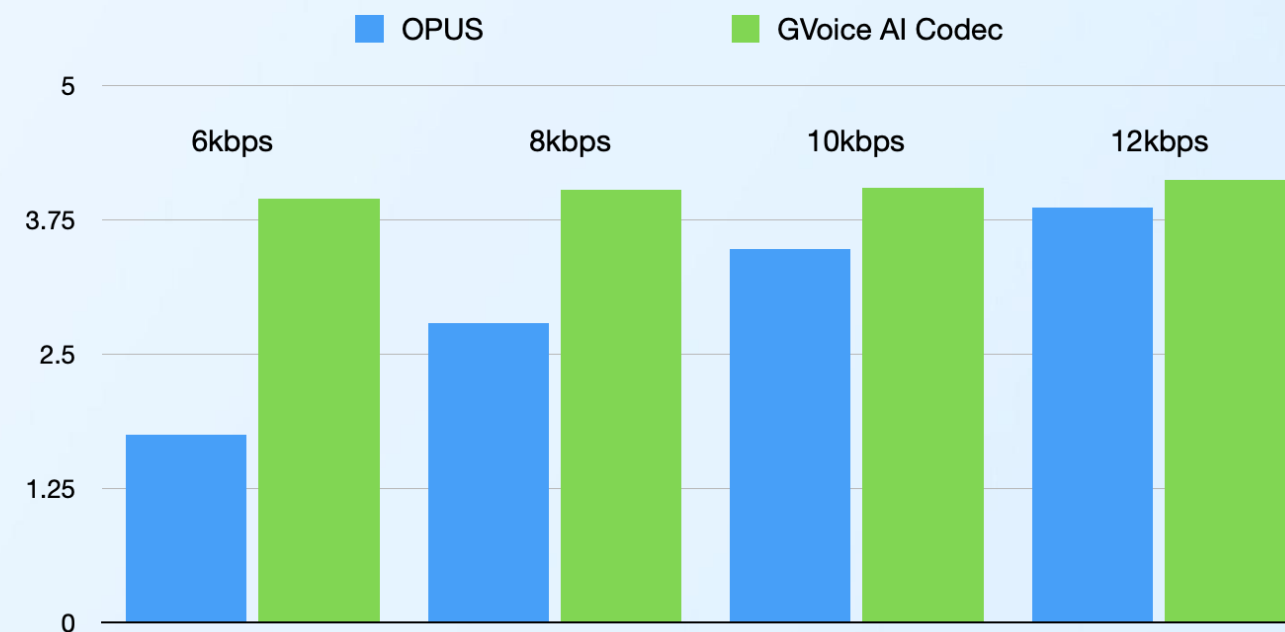
# Gvoice AI Codec: Major Components – Neural Packet Loss Concealment



- Inevitable packet loss due to weak connections, network speed fluctuation, etc
- Traditional Solutions: FEC, conventional PLC (zero insertion, waveform substitution or interpolation, etc)
- **Extra bit overhead and speech unnaturalness** in the portion of packet loss
- **Simulating the packet loss during model training:**
  - randomly masking the input encoded units
  - training the decoder to reconstructed masked portion from contextual frames/packets
  - no extra bit overhead and keep naturalness in speech
- Leading to good robustness up to **a 10% packet loss rate and a maximum continuous packet loss of 60ms**

# Gvoice AI Codec: Performances and Applications (1/2)

- Gvoice AI Codec has been fully adopted by some of the most influential mobile games, PEACEKEEPER ELITE (Game for Peace)
- Gvoice AI Codec at **5-12kbps** is able to compress and reconstruct speech with a quality comparable to that of 24kbps Opus.



- Compared to Opus, the speech quality (PESQ) improves significantly at all bit rates. The speech quality more than doubles under weak connection scenarios (e.g., 6kbps).





# Gvoice AI Codec: Performances and Applications (2/2)

- Comparisons with other open-source AI-powered speech codecs on benchmark evaluations (PESQ, POLQA, the higher the better)

Model	PESQ			POLQA		
	ITU	Meeting	HQ	ITU	Meeting	HQ
Lyra2-3.2k	2.18	2.83	2.29	2.55	3.1	3.04
Lyra2-9.2k	2.87	3.51	3.05	3.05	3.86	3.71
Encodec-6k	2.7	3.04	2.37	2.94	3.14	2.95
Encodec-12k	3.26	3.49	2.97	3.45	3.75	3.57
Encodec-24k	3.65	3.75	3.38	3.79	4.13	3.95
Gvoice AI Codec-10kbps	3.89	4.15	3.96	4.14	4.52	4.46

- We believe the successful application of Gvoice AI codec in Peacekeeper Elite has laid the foundation for broader use of AI-powered codecs, thus further improving the user experience for more games..

**PART.2**

## Haptic Feedback: How It Brings Inclusivity and Accessibility





# The Purpose of Deploying Haptic Feedback Technology



# Why do we care about haptic feedback

## Devices that require haptics



**Sense of reality**

Visible & Tangible

**Immersive experience**

Audiovisual, force-haptic multisensory stimulation

**Interactive experience**

Direct control of virtual objects, real-time sensory feedback

## Console Games

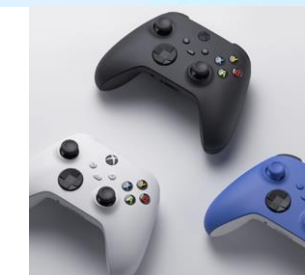
Haptic feedback is the core experience



Haptic feedback

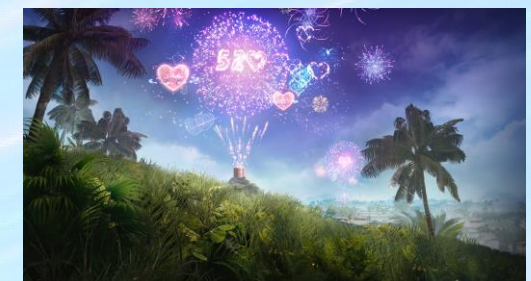


DualSense™ 无线控制器



## Mobile Games

Lack of feedback, apart from sound and visuals,





# Limitations on Mobile Devices

Diverse mobile devices and components make it difficult to achieve a consistent experience

Mobile phone model distribution among gamers



Mobile devices & Actuators

Bar ERM



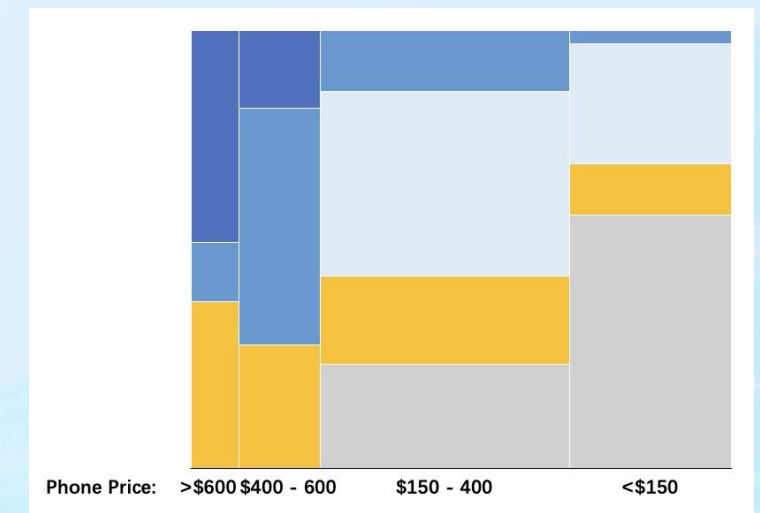
Coin ERM



Z-axis LRA



X-axis LRA



# Opportunities and challenges

## Opportunities

- Growing demand for more complex and immersive user experiences.
- The extensive use of linear resonant actuators (LRAs)
- Highly precise and responsive actuators for rendering vibrations

## Challenges

- The easy-to-use editing tool which can be adapted to the game engine
- The unified standard interface between application, operation system and hardware
- Rendering smooth and unified vibration effect



# PEACEKEEPER ELITE (Game for Peace)

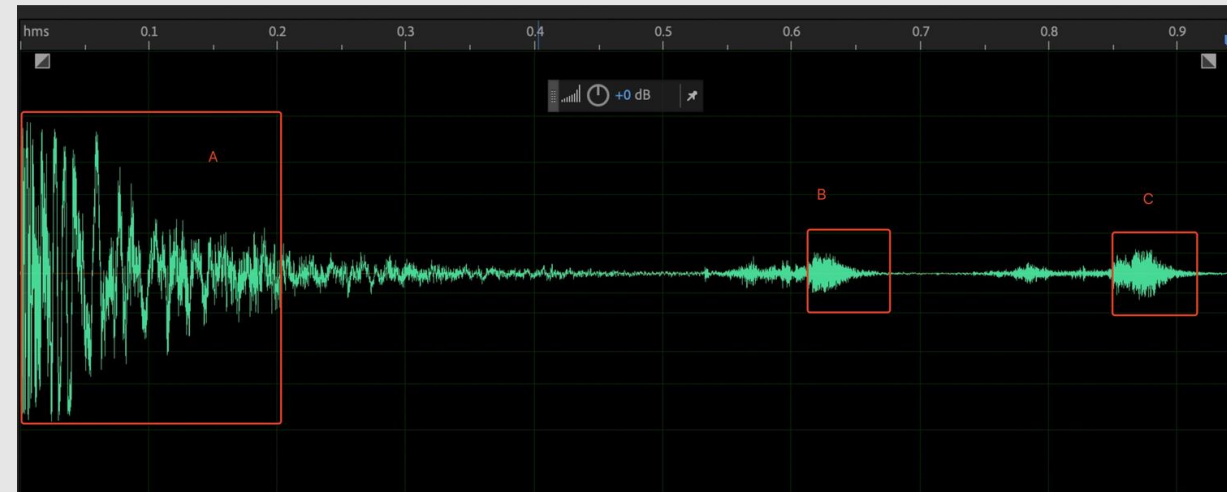
## Haptic Solutions





# How to design a haptic effect?

## Start designing a Haptic Effect



Sound wave of a shotgun from Peacekeeper Elite (Good for Peace)



### Independent design

Take some system vibrations built into mobile devices, such as incoming calls and notifications, as an example. Directly establish vibration rules to express the occurrence of specific events.



### Auditory-based design

Haptic Effect is designed to be used in scenarios that can be standalone, combined with auditory/audio, or combined with visual/animation

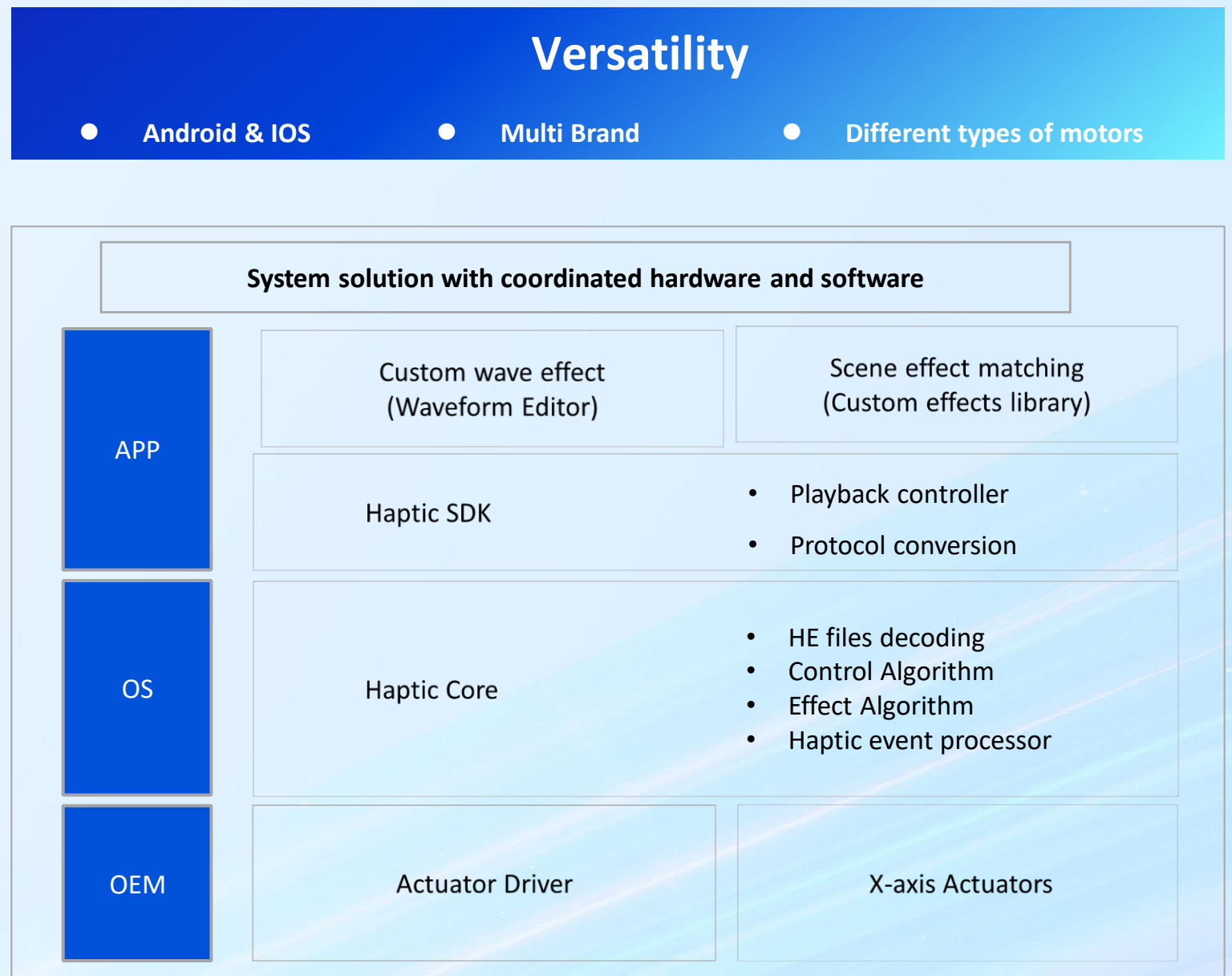
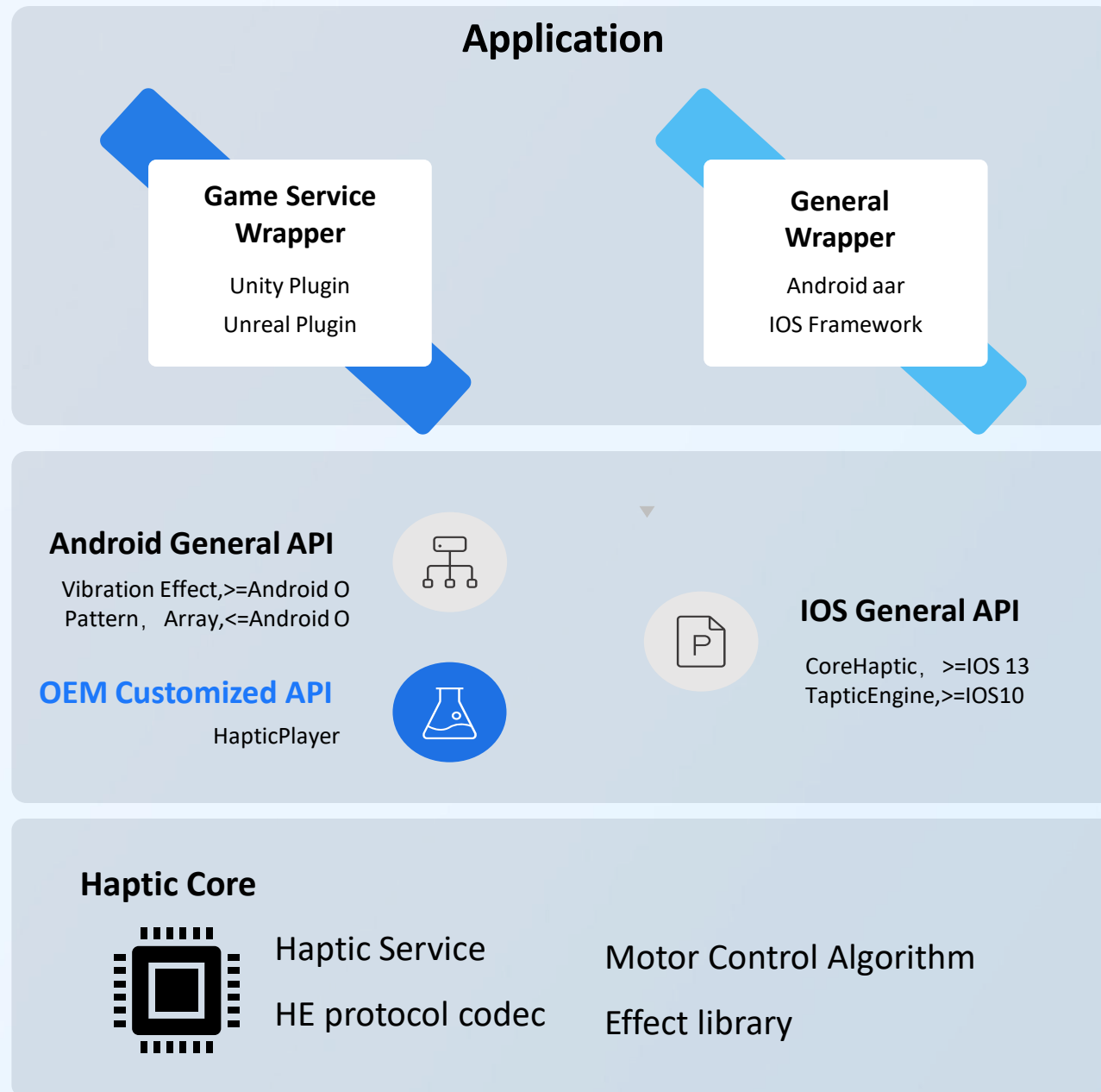


### Audio filtering to generate HE effect scheme

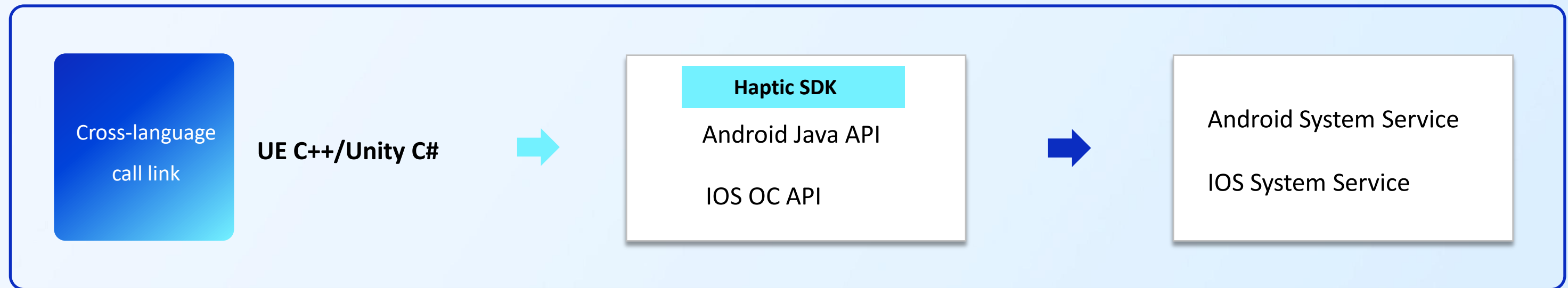
According to the audio waveform, the audio filter is transferred to the HE Event algorithm to generate the best HE description of the motor vibration.

# How the technical architecture of the haptic system is built?

Versatility and flexibility are the core design concepts throughout the entire solution



# How to optimize additional performance consumption?



## Optimization Solution

- 1 Asynchronous call for the time consuming tasks
- 2 Binary encoding of the haptic data, and reducing the data volume by 80%
- 3 Improving the caching for haptic data and decreasing the IO operation

## Performance test index

- 1 Time consumption of calls to the underlying vibration API
- 2 Memory consumption
- 3 HE impact on APP frame rate
- 4 Impact on device power consumption



# How to achieve synchronization of vibration, audio, and video design generation and playback?

## Vibration playback mechanism

01

### Static Mode

- ✓ read the vibration effect resources designed in advance at runtime all at once
- ✓ Audio, video and vibration play independently
- ✓ more suitable for shorter duration vibration effects
- ✓ Long video/audio design and playback defects

02

### Streaming Mode

- ✓ In the unit of vibration event, read and play according to the time stamp in HE
- ✓ Good for long vibration effects
- ✓ Synchronize with audio and video at the granularity of vibration Event

## Vibration Synchronization Scheme

01

### Synchronize with audio

- ✓ suitable for music player
- ✓ Audio segmentation, prefabricated multiple HE resources
- ✓ Synchronize by track, play regularly

02

### Synchronize based on video and animation effects

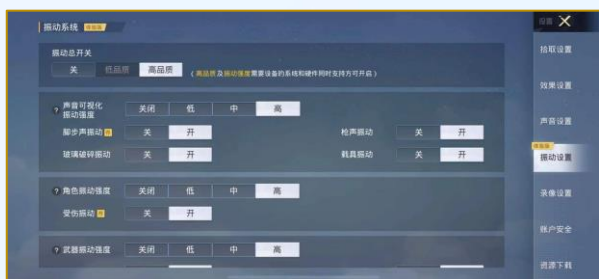
- ✓ Suitable for games, video display
- ✓ Video segmentation, prefabrication of multiple HE resources
- ✓ Timed playback according to video timestamp

# How does it work in games?

## Function



17 custom configurations



## Core elements



4 core elements, 200+ effects



firearms



vehicle

## Refinement effect



AWM



M416

## Application range

**5 million+**

Users open the function

**6k+**

models support

**200**

**million+**

devices support

# More than Gaming





# Helping visually impaired people experience the joy of gaming

In 2020, the World Health Organization estimates that 596 million people worldwide will be affected by visual impairment.

- **We talked to 100+ visually impaired users**

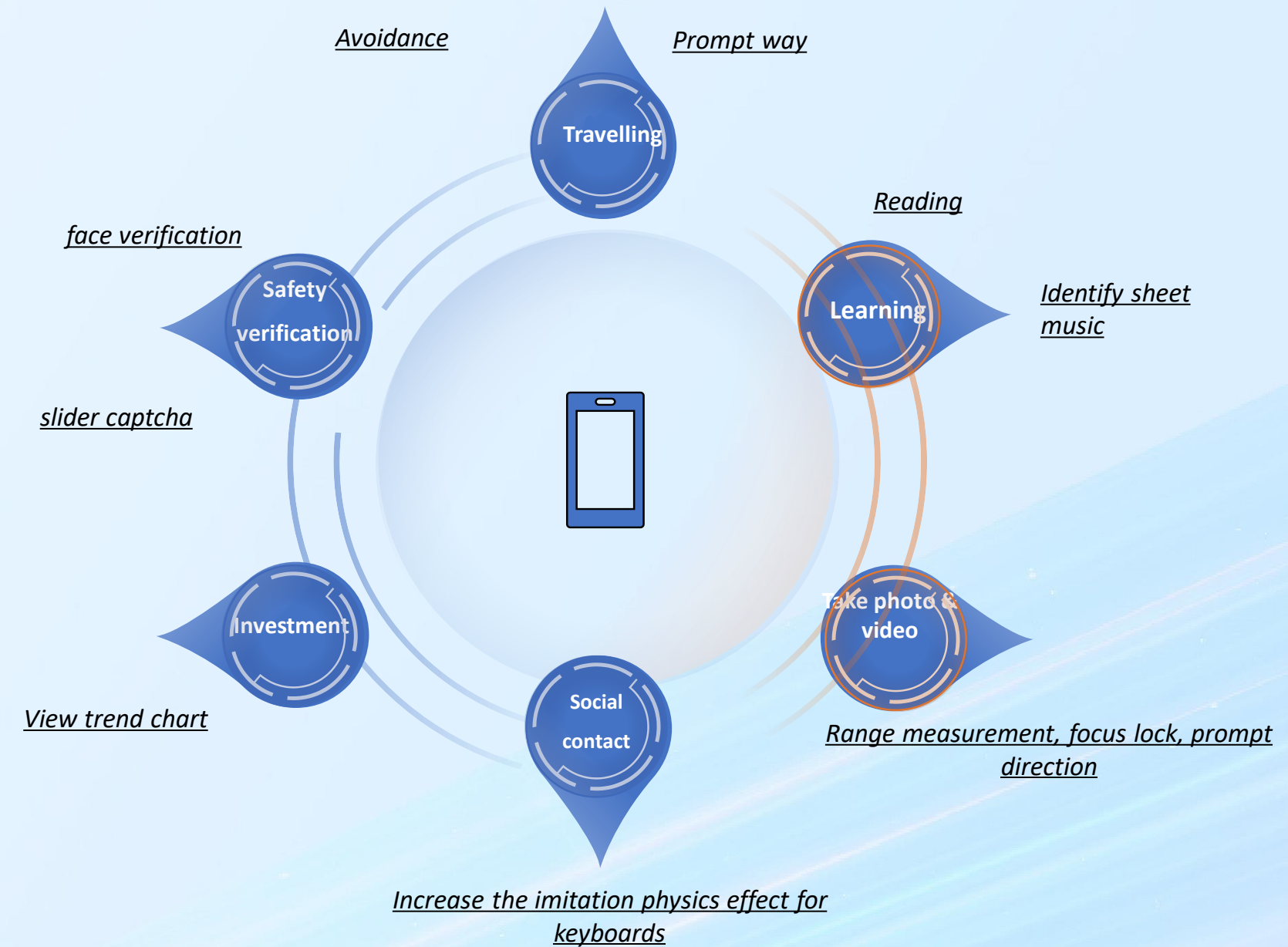
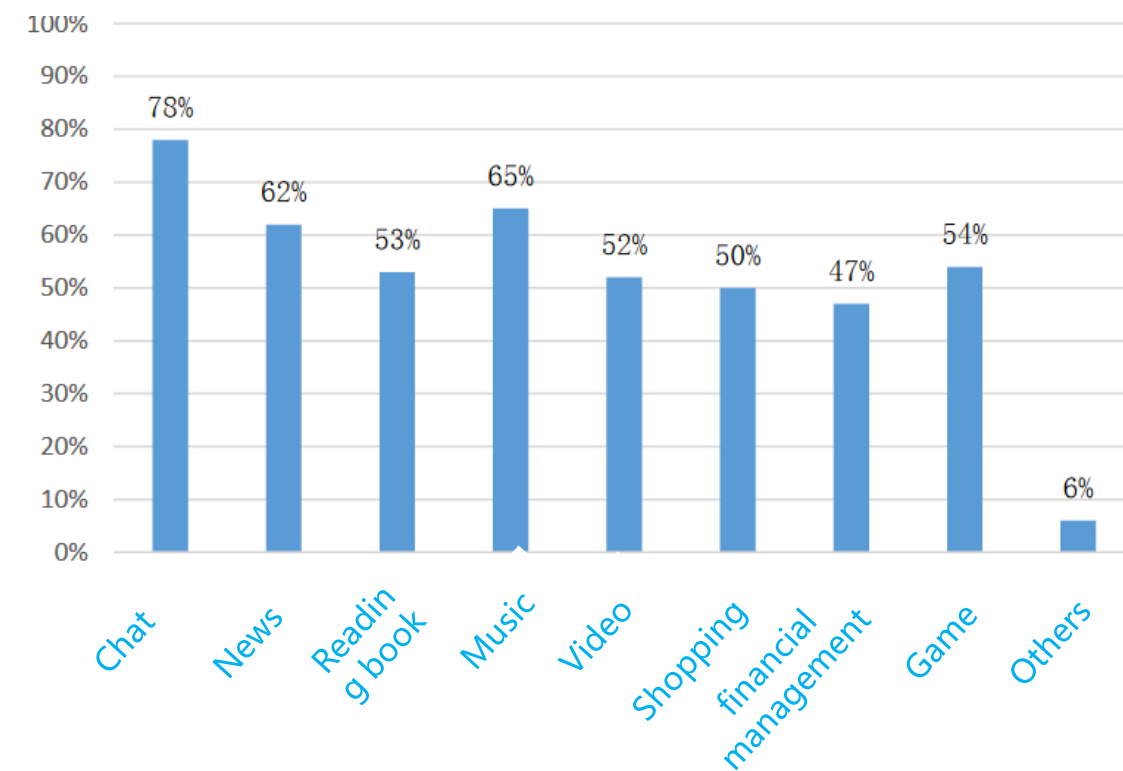


- **Haptics can help visually impaired users to gain more information and understanding.**

# Many scenarios that can be supplemented with information through vibration

## What do visually impaired people use their mobile phones for when accessing the internet?

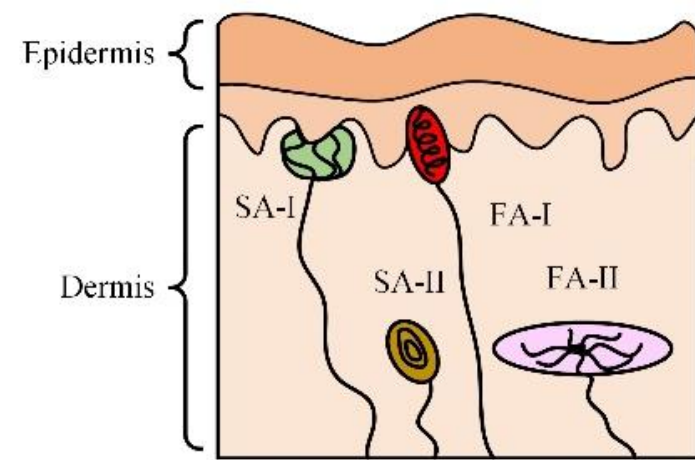
Here, we are surprised to see that this part of the user's information needs for the Internet encompasses a vast array of scenarios. Below is a list of their specific needs for travel, transactions, social media, and other areas.





# Get research support from universities in this field

When we sought to further implement the requirements, we also invited experts from universities to provide theoretical support. They assisted us in designing a more scientific approach for the visually impaired group in the field of haptics.



[1] SADATO N, PASCUAL-LEONE A, GRAFMAN J, et al. Activation of the primary visual cortex by Braille reading in blind subjects[J/OL]. Nature, 1996, 380(6574): 526-528. DOI:10.1038/380526a0.

Key point of Coding Design

**Location**

**Frequency**

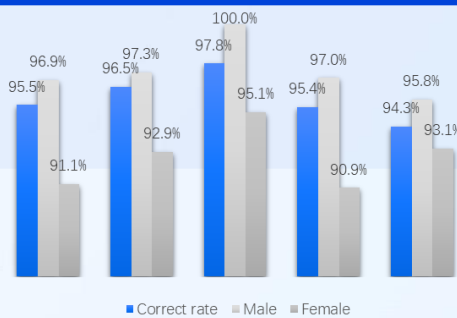
**Intensity**

**Duration**



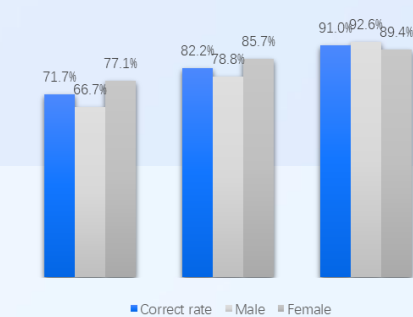
# Discover ways for visually impaired users to effectively extract information from three-dimensional sources

Taking intensity, frequency and duration as variables, several groups of experiments were conducted to determine the accuracy rate



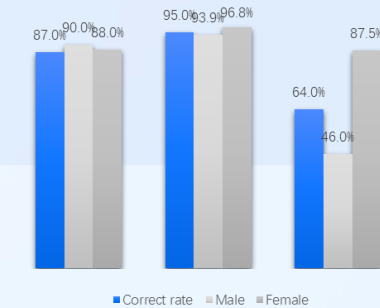
Take the intensity as the independent variable

tested intensity levels are 30 and 70.



Take the frequency as the independent variable

tested frequency levels are 40 and 70.



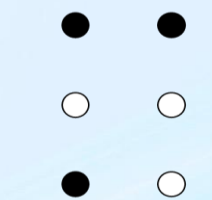
Take the duration as the independent variable

frequency level is 40, the intensity level is 70.

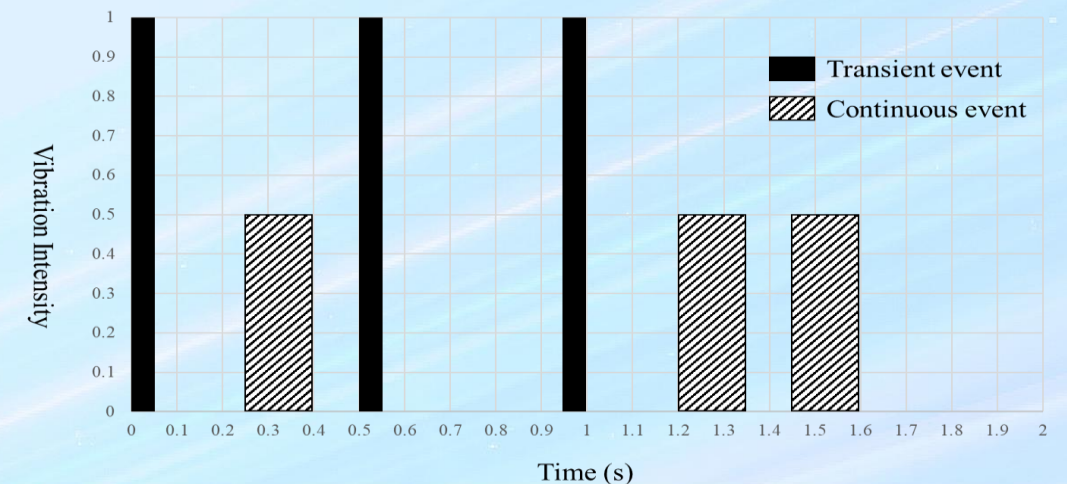
The compared results of the three are as follows

Variables	Highest correct rate
intensity	0.978
frequency	0.91
duration	0.95

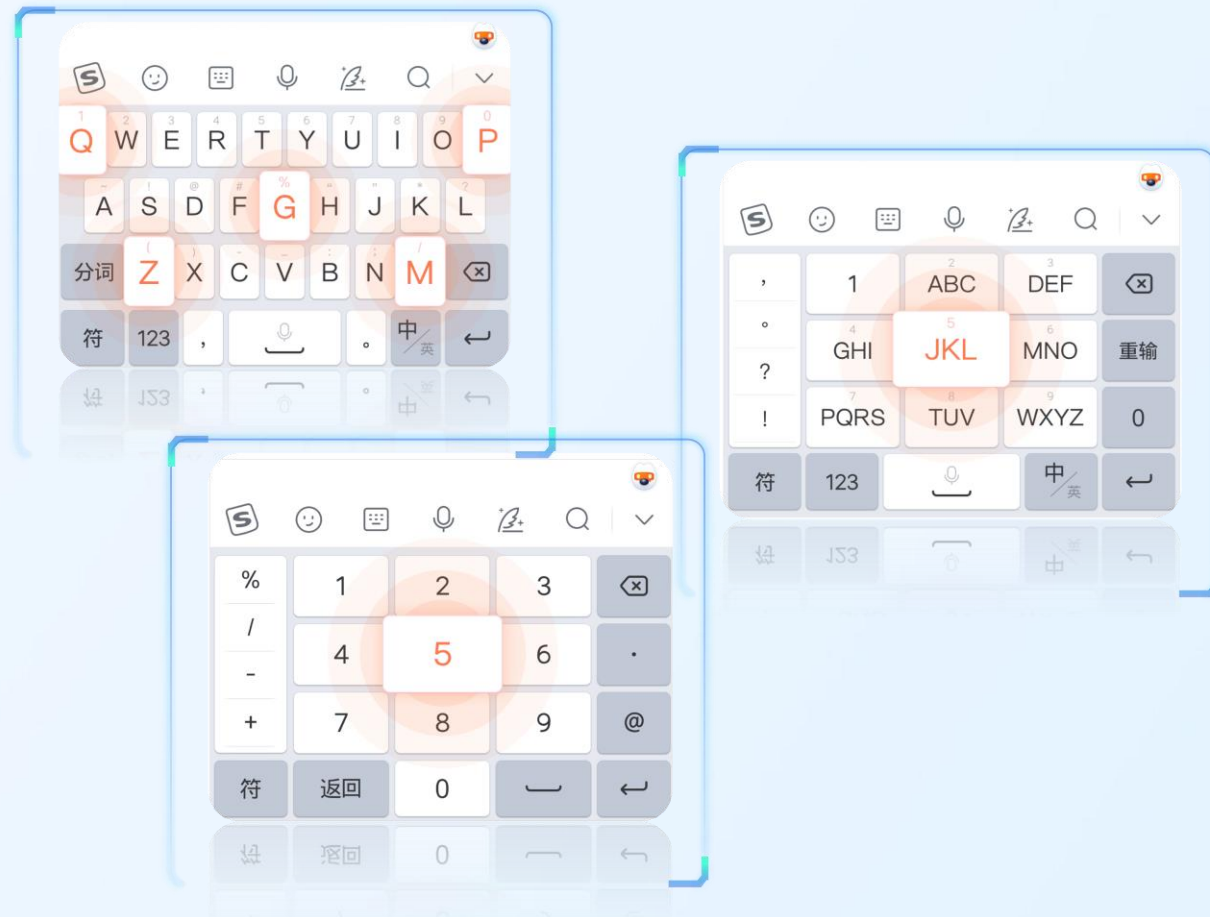
Coding methods of Six-point Braille



Braille character

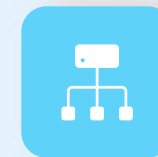


# In other basic application fields - input methods



## Inconvenient to use audio

Introduce different haptic effects for specific characters on the input method keyboard When audio feedback is inconvenient to use



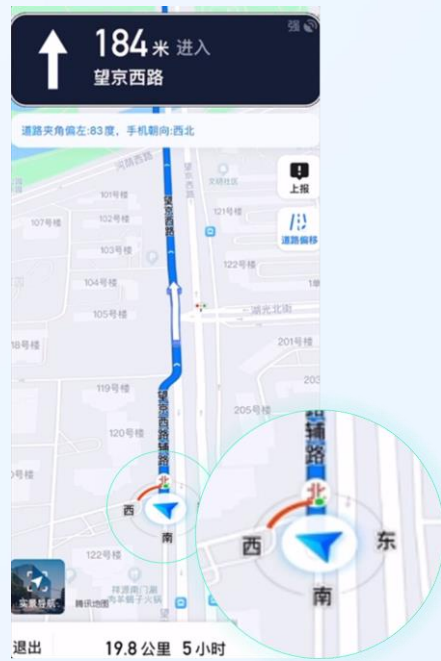
## Protect privacy

Recognize letters and numbers on the keyboard by haptics, while protecting user privacy to increase input efficiency and accuracy



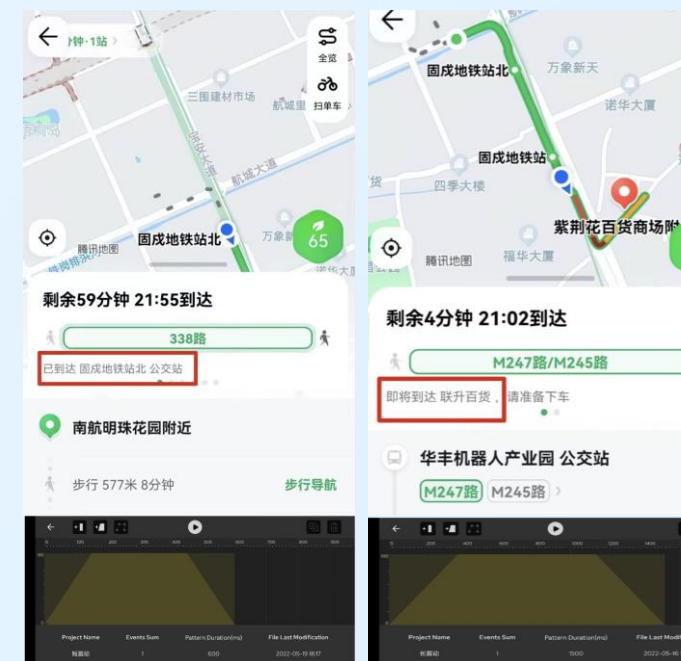
# In other basic application fields——navigation application

## Walk Navigation Accessibility Mode



When using walking navigation, if the user veers off the designated route, they will be alerted through vibrations and an audio announcement, guiding them back to the right direction.

## Transit Navigation Accessibility Mode



Through customized vibration effects, users are reminded in four different scenarios: waiting at the station, about to arrive at the station, arriving at the station, and arriving at the destination.



Let's see what users say

*Your day starts with opening your eyes*



# Global standardization efforts

Coordinate with global standardization organizations and bring harmonized game standards to industry  
Welcome to all interested parties to join our standard community



IEEE 2861 family  
Standards



## Global Standard Community

Initiate numerous IEEE working groups, head the team of more than 50 industry professionals;  
Participate in W3C, AVS, and Khronos initiatives;

## Establish Connection

Connect to other organizations, and open source community;  
Bring ideas from the entire sector;

## Build Consensus

Engage with industry stakeholders, solicit comments and seek agreement from all relevant parties;

## Collaborate with Academic

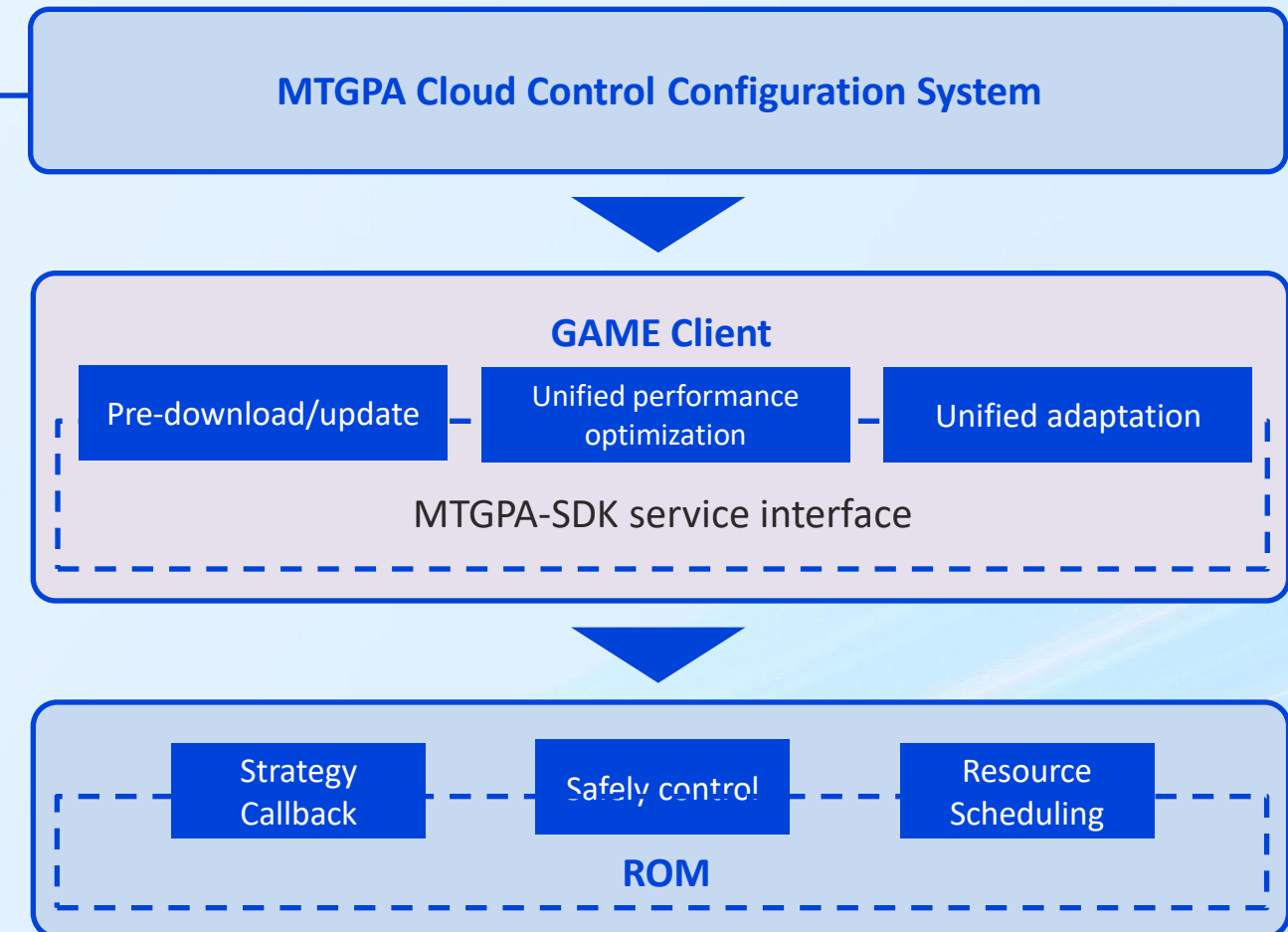
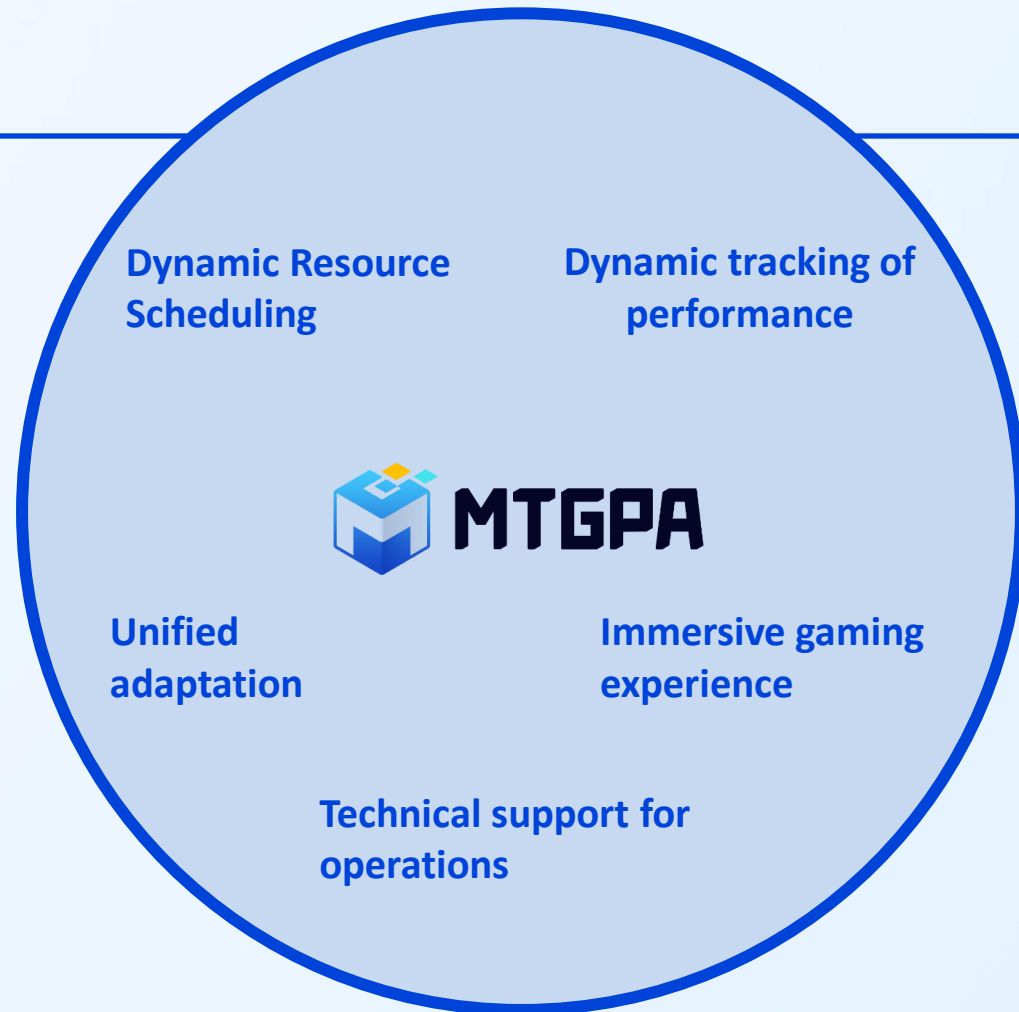
Partner with universities on exploring and innovating, focus on the objective of 'Technology for a good cause';

## Outcome

Publish paper on IEEE magazine, release announcement and white paper;

# MTGPA Includes multiple technical optimization capabilities and tools, welcome more Cooperation

## Helping studios with adoption of New technologies







# MEET LIGHTSPEED STUDIOS AT GDC2023

March 20-24, 2023 | San Francisco, CA