JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 29 No. 3, pp. 83-90, March 2024
https://doi.org/10.9708/jksci.2024.29.03.083

Design for Proximity Voice Chat System in Multimedia Environments

Jae-Woo Chang*, Jin-Woong Kim**, Soo Kyun Kim**

*Ceo, Braves. inc, Anyang, Korea
**Student, Department of Computer Engineering, Jeju National University, Jeju, Korea
**Professor, Department of Computer Engineering, Jeju National University, Jeju, Korea

[Abstract]

In this paper, we propose a solution to apply a proximity voice dialog system to voice dialog technology, one of the interaction systems in multimedia environments. A voice dialog between multiple users in a multimedia space is designed by adjusting the volume of the voice according to the distance between the user avatars and muting the user who is beyond the audible distance. The main feature of this research is a reliable UDP-based active server system that delivers low-quality voice data to users who are far away based on distance and does not transmit voice data to users who enter the inaudible area for economic development. The performance of the proposed system was measured in a previously completed project based on the Unity game engine, and it is expected that the system proposed in this research will be actively used in environments that provide interaction between multiple users such as met averse content and real-time battle action games.

▸ **Key words**: Proximity Voice Chat, Multimedia, Virtual World, reliable UDP, Unity Game Engine

[요 약]

본 연구에서는 멀티미디어 환경에서 상호작용 시스템 중 하나인, 음성 대화 기술에 대하여 근접 음성 대화 시스템을 적용하는 솔루션을 제안한다. 사용자 아바타들 간 거리에 따라 음성의 볼륨을 조절하고, 가청 거리를 벗어난 사용자에게는 음소거를 적용하는 방식으로 멀티미디어 공간에서 여러 사용자 간의 음성 대화 방식을 설계하였다. 본 연구의 가장 큰 특징은 경제적인 개발을 위해, 거리를 기반으로 먼 거리에 있는 사용자에게는 저음질의 음성을 전달하고, 비 가청 지역에 들어선 사용자에게는 음성 데이터를 전송하지 않게 하는, reliable UDP 기반 능동적 서버 시스템에 있다. 제안 시스템은 사전에 완성하였던 유니티 게임 엔진 기반 프로젝트에서 성능을 측정하였으며, 본 연구에서 제안한 시스템을 메타버스 콘텐츠, 실시간 대전 액션 게임과 같이 여러 사용자 간 상호작용을 제공하는 환경에서 적극적으로 이용되는 것을 기대할 수 있다.

▸ **주제어**: 근접 음성 대화, 멀티미디어, 가상세계, reliable UDP, 유니티 게임 엔진

• First Author: Jae-Woo Chang, Corresponding Author: Soo Kyun Kim
*Jae-Woo Chang (ami@bravegames.co.kr), Braves. inc
**Jin-Woong Kim (0802dragon@naver.com), Department of Computer Engineering, Jeju National University
**Soo Kyun Kim (kimsk@jejunu.ac.kr), Department of Computer Engineering, Jeju National University
• Received: 2024. 02. 27, Revised: 2024. 03. 21, Accepted: 2024. 03. 22.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

Opus Codec[12]과 같이 실시간 음성 대화에 관한 연구는 게임이나 여러 멀티미디어 환경에서의 필요성으로 인해 발전하고 있는 분야이다. 그중에서도 근접 음성 대화 서버 솔루션 개발에 관한 연구는 기존 대형 플랫폼 의존도를 줄이고, 소규모 회사도 경제적으로 음성 대화 기능을 자체적으로 구축하는 방안으로 중점을 두게 되었다. 본 연구는 기존 연구[2]와 같이 과도한 비용과 데이터 전송의 비경제성 문제를 해결하기 위해 진행되었다. '어몽어스(Among Us)' 게임의 근접 음성 대화 기능[3]에서 영감을 받아, 사용자들이 게임 내에서 서로 가까이 있을 때만 음성으로 소통할 수 있는 시스템을 개발하는 것을 목적으로 한다. 이러한 기능은 물리적 거리에 따라 음성의 볼륨을 조절하여 현실감 있는 상호작용을 제공한다[3] 그리고 네트워크 서버 구성, 데이터 전송 최적화, 오디오 코덱 적용이나 사용자 간 인터랙티브 음성 대화 구현에 대한 다양한 기술적 접근 방식이 제시되었다[4, 5].

본 논문에서는 근접 음성 대화 시스템의 개발 과정, 기본 구성 방식, 중요 알고리즘, 그리고 기능 테스트 결과를 보여준다[6, 7, 8]. 특히. 음성 데이터의 전송률을 획기적으로 줄이면서도 사용자 간 음성 대화 기능을 극대화하는 방법을 탐색, 경제적이고 효율적인 근접 음성 대화 시스템의 구축 하였다. 이러한 기술은 메타버스 콘텐츠, 잠입 대전 FPS 게임, 가상공간 공연, 가상 오피스 리모트 회의 및 지도 내 사용자 간 약속 및 상대 찾기 등 다양한 분야에 적용될 가능성을 보여준다.

II. Preliminaries

1. Related works

근접 음성 대화 서버 솔루션 개발에 관한 최근 연구 동향은 대형 플랫폼 서비스를 대체할 수 있는 경제적이고, 효율적인 솔루션의 필요성에 초점을 맞추고 있다[9]. 특히, '어몽어스(Among Us)' 게임에서 채택된 근접 음성 대화 기능은 사용자 간의 실시간 음성 소통 방식에 혁신을 가져왔으며, 이를 통해 메타버스 내에서 현실감 있는 상호작용을 제공하는 새로운 가능성을 열었다[10]. 본 연구는 이러한 기술적 접근을 활용하여 음성 데이터 전송률을 최적화하고, 사용자 간의 인터랙티브한 음성 대화를 구현하는 방법을 연구하였다. WebRTC와 같은 기존 P2P 기반 솔루션은 데이터의 불안정성과 보안 취약성 등의 문제로 인해 범용적인 사용에 한계가 있다[11]. 이에 대한 해결책으로, 본 연구는 경제적이며 효율적인 근접 음성 대화 시스템의 개발을 통해, 다양한 온라인 플랫폼과 게임에서 실시간으로 현실감 있는 음성 대화를 가능하게 하는 새로운 방안을 제시한다. 또한, 이 시스템은 소규모 회사나 개발자들이 대형 트래픽으로 인한 과도한 비용 부담 없이 자체 음성 대화 기능을 구축할 수 있게 함으로써, 음성 대화 기반 서비스 개발의 새로운 장을 열고 있다.

III. The Proposed Scheme

근접 음성 대화 시스템을 위한 알고리즘은 다음과 같이, 네트워크 서버 구성, 데이터 전송 방식 결정, 그리고 코덱 적용의 세 가지 주요 구성 요소로 분류할 수 있다. 본 장에서는 구현 방법을 소개한다.

1. Multimedia Server Developing Environment

음성 대화 혹은 화상 대화는 줌(Zoom), 구글의 미트(Meet), MS의 팀즈(Teams) 등 대형 플랫폼 서비스를 통해 쉽게 접해볼 수 있는 기능 중 하나이다. 그러나 정작 이 기능을 자체적으로 구축해보고자 할 때 소규모 회사가 감당해야 할 비용이 너무 과하여 접근이 쉽지 않은 영역이기도 하다.

음성 대화 및 서버 운용에 관련된 기술 확보는 차치하더라도 온라인망을 통해 음성, 영상 데이터가 이동해야 한다는 기본 조건이 있는 한, 대형 트래픽에 대한 크나큰 비용은 피해 갈 수 없다는 비일반성과 비경제성을 내포하고 있기 때문이다.

물론, 대형 트래픽을 회피해 갈 방안으로 webRTC와 같이 P2P 환경을 기반으로 한 솔루션도 제안되고 있으나 데이터의 불안정성, 보안 취약, 웹브라우저를 사용하지 않는 환경에서의 기능 제약, 대규모 사용자들 관리 취약 등으로 인해 이런 솔루션의 사용은 범용적일 수 없다는 단점을 가지고 있다. 최근 들어 마피아 게임의 한 종류인 '어몽어스(Among Us)'라는 게임 콘텐츠가 글로벌한 성공을 거두면서 그 기능을 좀 더 인터랙티한 환경에서 즐기기 위해 근접 음성 대화(Proximity Voice Chat)라는 컨셉이 등장하며 많은 관심을 불러일으키고 있다. 일부 사용자들과 몰래 만나 그들과만 대화하고자 할 때, 메신저 타이핑이 아닌, 직접 일부 사용자들끼리만 음성 대화를 할 수 있는 환경에 적합한 기능이었기 때문이다. 또한, 이러한 컨셉은 메타버스 내에 현실감 넘치는 세계 구축이라는 명제를 해결하는

방안으로 관심을 불러일으키고 있다. 이에 음성 및 영상에 관련된 기술 중, 인터랙티브한 환경에서 활용하는 음성으로 관심의 폭을 좁힐 경우 근접 음성 대화를 통한 다양한 적용이 가능할 뿐 아니라 새로운 기술 접근을 통해 경제성을 띤 개발 및 서비스를 할 수 있음을 확인할 수 있다. Fig. 1은 근접 음성 대화 시스템의 기본 구현 콘셉트를 나타낸다.

2. Proximity Voice Chat System

지금까지 음성 대화는 가상 세계의 한 방 안에 있는 모든 사람이 같은 소리를 같은 볼륨으로 듣는다는 것을 기본으로 해서 서비스됐다. 반면 근접 음성 대화는 현실의 물리 세계를 바탕으로 나의 아바타와 가까이 있는 상대에게는 소리가 또렷하게 전달되나 거리가 멀어질수록 전달되는 소리의 볼륨이 줄어들고, 일정 거리를 벗어나면 비가청 지역이 된다는 특징을 가지고 있다. 즉, 사용자들이 본인의 의지로 상대와의 거리를 달리함에 따라 소리의 전달 유무가 갈리게 되는 점을 통해 더 적극적인 온라인 인터랙션을 이끌어내는 것이 바로 이 근접 음성 대화의 특징이다.

근접 음성 대화를 구현하는 방법은 개발 방식에 따라 다양한 형식을 띨 수 있다. 음성 데이터 전달에 대한 비용적 부담을 무시할 경우, 그리고 개발에 공수를 많이 추가하고 싶지 않을 경우, 일반 음성 대화 방식과 동일하게 개발한 후 사용자 아바타들 간 거리에 따라 소리의 볼륨을 거리별로 조절해 주고, 가청 거리를 벗어난 사용자에게 전달된 음성 데이터는 소리 끄기(Mute) 기능으로 소리를 끔으로써 들리지 않도록 할 수 있다.

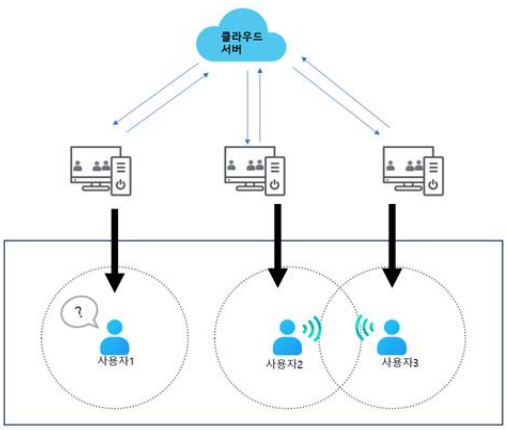

Fig. 1. Basic Implementation Concept for Proximity Voice Chat

한편, 근접한 상황에서만 음성 대화가 가능하게 하는 상황 개발을 좀 더 적극적으로 구성하면 음성 대화 서비스 부분에 비용적으로 경제성을 띤 개발이 가능하다는 가능성을 확인하였다. 본 개발의 가장 중요한 방향성은 거리를 기반으로 해서 거리가 멀어져 간 사용자에게는 저음질의 음성을 전달하고, 비 가청 지역에 들어선 사용자에게는 아예 데이터를 전송하지 않도록 서버 시스템을 능동적으로 구성하는 것이다. 즉, 거리를 기반으로 전송하는 음성 데이터의 크기와 전송 여부를 조절함으로써 지금까지 일반화되어 왔던 전송과 달리, 음성 데이터들의 수량과 크기가 획기적으로 줄어듦에 따라 소규모 음성 서비스를 꾸리고자 하는 소기업 또한 이 기능을 통한 경제적인 서비스가 가능하다는 것이다.

3. Basic Configuration

거리를 기반으로 소리가 들리고 안 들리는 차별점을 활용하되 서버를 기반으로 전송할 데이터의 성격을 미리 구분하여 전송 여부를 결정하는 만큼 네트워크 활용 방식에 대해 좀 더 명확한 활용 영역의 구분이 필요하였다.

3.1 Network Server Configuration

서버의 기본 구성은 표준화된 서버와 클라이언트가 연계된 통신 모델을 기반으로 하기 위한 한 일환으로 클라우드서비스를 활용하였으며 여기에 가상 세계에서 움직이고 있는 아바타들의 위치 정보 데이터를 기반으로 거리별 최적화 데이터를 전송하는 알고리즘을 포함했다. 여기에 데이터 전송을 위한 네트워크 방식 결정에 대한 다양한 테스트가 병행되었다.

클라우드 서버를 통해 왕복 되는 데이터 중 가장 주를 이루는 것은 ① 아바타 위치 정보, ② 콘트롤 데이터, ③ 음성 데이터로 구분된다.

아바타 위치 정보는 음성 대화와 직접적인 연관성은 없으나 서버에서 클라이언트로 전송할 음성 데이터 성격을 선정할 중요 정보가 된다. 서버로 전송된 사용자의 음성 데이터는 거리 정보를 기반으로 해서 고음질, 저음질 데이터 중 어느 데이터를 전송할 것인지, 혹은 데이터를 전송할 것인지 전송하지 않을 것인지 서버에서 판정하게 되는데 이러한 결정을 수행하게 되는 판단 척도가 바로 아바타의 위치 정보이다.

본 정보를 통해 서버가 일정한 판정을 하는 절차가 있는 만큼, 일반 서버 작업과 비교할 때 추가 연산이 필요한 단점이 생길 것이나 클라우드 서버를 활용하여 연산을 진행하고 음성 데이터를 전송하는 과정을 전체적으로 볼 때, 전송하게 될 데이터들이 획기적으로 줄어들게 되는 만큼 효율을 위한 작업이라고 할 수 있다.

컨트롤 데이터는 일반적인 대화 상황에서 특정 사용자가 하는 말을 듣고 싶지 않을 경우, 소리 끄기 요청 등 클라이언트에서 서버로 발송하게 될 요청 정보들의 한 종류가 될 수 있다. 이러한 정보들은 사용자가 직접적으로 요청을 하는 만큼 정확하게 수행할 필요성이 있는 데이터들이다. 반면 음성 데이터는 가상 세계에 모여있는 많은 사람이 서버로 전송하게 되는 대화 관련 데이터 덩어리들이다. 이 데이터들은 다른 클라이언트로 전달되어 음성으로 다시 출력될 필요가 있는 데이터들인 만큼 화자가 말하는 순간 가장 빠르게 전달될 필요성이 있는 데이터들이다.

클라이언트와 서버 사이의 컨트롤 데이터를 안정적으로 전달하기 위해 TCP(Transmission Control Protocol : 전송제어프로토콜)를 기본적으로 고려할 수 있다. 본 제어 프로토콜은 속도는 느리지만 높은 신뢰성을 보장하는 만큼 콘트롤 데이터 전송에 적합한 형식이며 이를 통해 전송 도중 일어날 수 있는 전송 오류를 자연스럽게 복구하는 절차를 이용, 안정적인 데이터 전송을 완료할 수 있는 방향성을 검토하였다.

음성 데이터 전송의 경우 가장 빠른 속도로 고음질을 전달해야 하는 것이 최우선인 만큼 UDP(User Datagram Protocol : 사용자 데이터 그램 프로토콜)를 활용한 통신 채널을 통한 전송을 고려하였다. UDP는 비연결형 서비스이기 때문에 정보를 주고받는 신호 절차가 존재하지 않으며 흐름 제어 혹은 혼잡 제어와 같은 기능은 처리하지 않아 신뢰성 있는 데이터 전송은 보장하지 못하지만, 정보를 빠르게 전달하고 네트워크 부하가 적은 장점이 있다.

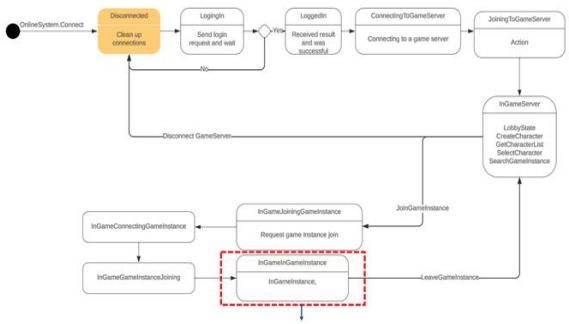

Fig. 2. Apply voice channels to in-game instances. Establish an environment where all voice-related data can be exchanged using a single channel

음성정보 데이터는 일부가 전송에 실패하여 최소의 보완 기능 적용, 혹은 복구 처리 없는 음성 출력을 수행하더라도 0.1초 미만으로 매우 짧은 내용인 만큼, 사용자는 빠진 것이 있다는 것을 거의 인지하지 못하거나, 듣고 이해하는 데 큰 문제가 발생하지 않으므로 신뢰성보다는 연속성이 중요한 데이터 영역이라 할 수 있다.

컨트롤 데이터 역시 전달 안정성이 중요하지만, 음성 데이터는 빠른 연속성이 중요한 영역의 데이터들이다. 초기에는 TCP와 UDP를 채널별로 분리하여 적용하는 2채널 방식을 검토하였으나 최종적으로 2가지 요구 조건을 동시에 만족시킬 수 있는 rUDP(reliable UDP)[11]를 활용한 단일 채널 방식의 네트워크 구성을 완성하였다. 이를 통해 안정성과 빠른 연속성을 통시에 만족시킬 수 있는 환경을 구축할 수 있게 되었다. Fig. 2는 게임 내 개체에 대해 음성 채널을 적용하는 단계를 설명한다.

3.2 Applying Codec

오디오 파일을 전송할 때 디지털 오디오를 압축하고 압축을 해제하는 알고리즘을 가지고 있는 오디오 코덱을 통해 데이터 크기를 좀 더 작은 크기로 만드는 것이 일반적이다. 이에 CPU 사용량은 더 높지만 Vorbis보다 좋은 음질로 더 많이 압축할 수 있으며, 대사같이 반복 재생되지 않거나 환경음과 같이 긴 오디오 음원에 적합하여 음성정보 전송에 가장 유리하다고 알려진 공개 소스 코덱인 오퍼스(Opus)[12]를 활용하여 시스템 라이브러리를 구축하였다. 오퍼스는 지연 시간이 아주 낮으면서도 고음질을 목표로 만들어진 CELT 코덱과 스카이프(Skype)의 고효율 음성용 코덱으로 만들어진 SILK를 통합하여 32kbps에서 가청 주파수를 온전히 커버하는(Fullband, 최대 20kHz) 고음질 음성 코덱이다.

소리는 파동의 형태로 이동하며, 이러한 파동은 주파수가 서로 다른 진동으로 인해 발생한다. 이러한 주파수는 Hz 단위로 측정되며, 우리는 주파수를 음의 높이로 인식한다. 인간의 일반 음성은 80Hz~14,000Hz 범위인 것으로 알려져 있으며 이를 기준으로 해서 표준 전화 통신의 경우 보통 8kHz 또는 16kHz를 택하고 있다.

오퍼스 코덱은 실시간 인터넷 음성 통신과 같은 저 비트레이트 상황에서 극한의 효율성을 추구하도록 개발된 포맷이기 때문에, 풀밴드(Fullband)의 경우라도 20kHz 이상의 신호는 기록하지 않고 있으며 이 신호를 최종적으로 48kHz로 변환하여 사용한다.

오퍼스 코덱의 특성 상 CD의 주파수인 44.1kHz가 없으며 8, 12, 16, 24, 48과 같은 배열로 코덱의 단순화를 꾀하고 있다. 특히 20kHz 이상의 기록 주파수를 사용하고 있지 않으며 20kHz의 신호는 샘플링 주파수 48kHz로 리샘플링하는 특성을 보여주고 있다. 이러한 코덱의 특성을 고려하고 사람의 음성을 최대한 간결하게 전달할 필요성을 만족시키기 위해 처음 전송에 설정되는 음성의 음질은

16bit/24kHz로 설정하였다. 이 음질은 기본적으로 384kbps의 전송률을 가지게 된다.

근접 음성 대화 시스템을 구축하기 위해 먼저 클라이언트 내에 Opus 코덱이 적용된 라이브러리를 연동하고 화자를 통해 음성이 출력될 때 2개의 음질로 인코딩되는 절차를 거치는 작업을 수행한다. 미리 설정한 16bit/24kHz로 사용자 음성이 마이크를 통해 클라이언트 코덱에 전달되고 여기서 소리의 압축이 이루어지게 된다. 1번째 음질은 가장 가까이 있는 사용자들과의 대화를 전달하기 위해 풀 밴드의 주파수를 사용하여 20kHz로 했으며 먼 거리에 있는 사용자에게 전달되는 볼륨이 줄어든 소리의 경우는 고음질로 전달이 필요 없는 만큼, 고음질의 30% 크기로 줄어드는 용량만큼의 압축률을 사용하였다. 일단 음성 대화의 경우 고음질 원본 1개만 서버로 전송하는 시스템이지만 최적의 근접 음성 대화 시스템 구축 및 데이터 전송의 최적화를 위해 2개의 음질을 인코딩하여 전송을 준비하는 것이 가장 유리하다는 결과를 확인하였기 때문이다. 그림 3은 아바타 간 거리에 따른 음질의 변화를 나타낸다.

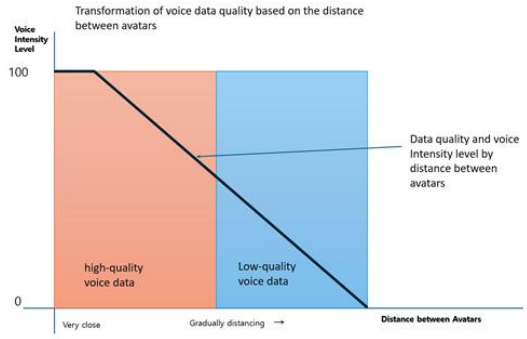

Fig. 3. Graph expressing the correlation between sound quality and volume applied according to the distance between avatars

3.3 Function Test

본 근접 음성 대화 시스템의 구축 및 테스트를 위해 미리 완성하여 보유하고 있던 Fishingonline.exe 메타버스 클라이언트를 활용하였다. 본 시스템은 인텔 i9-11세대를 사용하고, 그래픽카드 지포스(GeForce) GTX 3080, 12GB를 사용하고 메인메모리는 32GB였으며 유니티 2022.3.17f버전과 C#4.0이 사용되었다. 본 클라이언트 활용의 편의를 위해 디버깅모드를 사용하여 사용자 간 거리를 측정하는 기능을 추가하였으며 음성 대화가 시작되는 시점에 관련, 채널이 열리는 정보, 전송된 데이터의 사이즈 및 시간을 확인할 수 있는 로그 파일을 준비하여 기록하였다.

본 테스트를 통해 가상 세계 안의 아바타 간 거리에 따라 음성 볼륨이 조절되고 서버 설정값을 통해 미리 지정해 놓은 20m 이상 멀어지게 되면 데이터 전송이 막히며 소리가 들리지 않는 것까지 확인할 수 있었다. 한편 3차원 가상공간 내에 아바타들이 움직이며 서로에게 하는 이야기를 듣게 되는 환경인 만큼 자연스럽게 상대가 얘기하는 음성의 방향성이 필요한, 환경 설정 과정이 필요하였다. 음성의 방향성을 설정하기 위해 기본적인 음성은 모노로 인코딩하였다. 그리고 사용자 간 위치 방향성에 따라 3D의 방향성을 갖게 하도록, 간단히 설치 및 설정을 할 수 있는 3D 오디오 API OpenAL 1.1을 활용하였다.

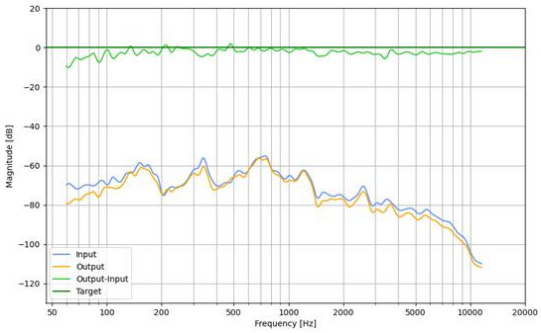

Fig. 4. Compare the wavelength when the voice is input and the wavelength of the voice output after passing through the server and reaching another client

아래 테이블 1은 2명의 아바타가 서로 1m 이내의 거리에서 대화하고 있을 때 뽑은 로그이다. 첫 번째 줄을 보면 서버로 보내지는 파일은 427, 서버에서 받는 파일은 339이다. 처음 보내질 때 고음질 339와 저음질 88이 동시에 보내지면서 두 개를 합한 427의 크기로 서버에 전송이 되고, 서버는 거리를 판단한 후 고음질 339만 전송한 것을 볼 수 있다.

Table 1. Size and Distance for Voice Data

Voice Data Size(byte)	High-Quality Voice Data	Low-Quality Voice Data	Data size from far distance compared to near distance
427	339	88	0.259587021
384	301	83	0.275747508
398	305	93	0.304918033
425	338	87	0.25739645
427	335	92	0.274626866
427	341	86	0.252199413
406	322	84	0.260869565
424	340	84	0.247058824
427	338	89	0.263313609
418	331	87	0.262839879
425	341	84	0.246334311
429	338	91	0.269230769

전송되었던 음성 데이터 정보와 전송받은 데이터의 양을 기본으로 하여 고음질 데이터의 크기, 저음질 데이터의 크기, 그리고 고음질과 저음질 데이터의 크기 비례를 확인할 수 있었다. 각 상황에 따라 음성의 압축률과 데이터 크기 등에 약간의 변동은 있을 수 있으나 저음질은 고음질 대비 약 26.5% 크기에 해당하는 것을 확인할 수 있었으며 이 음질은 와이드 밴드(Wideband)에 해당하는 수준이다.

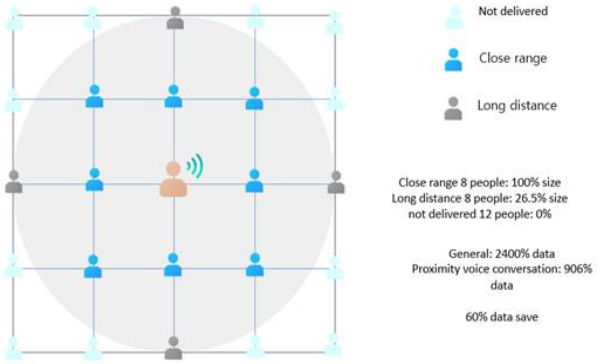

Fig. 5. Proximity Voice Chat

가상 세계에 사용자들의 배치와 환경에 따라 다양한 형태로 변동이 가능한 상황이나, 단순히 4각형 형태의 필드에 동일한 거리 간격을 가지고 사용자들을 배치하고, 그림에 보이는 회색 영역에 있는 아바타에게 음성 전달이 가능한 환경을 가정해 볼 경우, 근거리 위치에 있는 아바타는 8명, 원거리에 위치하는 아바타는 4명이다. 이때 음성이 아예 전달될 수 없는 위치에 있는 아바타는 12명. 전체 계산을 해볼 경우, 화자가 발송한 데이터 크기가 100이라 하면 906의 데이터가 서버를 통해 전송됨을 알 수 있다. 반면, 근접 음성 대화를 적용하지 않는 일반 환경의 경우 화자를 제외한 24명에게 동일한 데이터가 전송되는 만큼 2400의 데이터가 전송되는 것이라 할 수 있다. 이 2개의 데이터를 비교하여 906의 데이터가 2400의 데이터보다 2.6배의 효율성을 보여주는 것을 확인할 수 있다. 위 그림은 단순한 한 예에 불과하나 근접 음성 대화가 일반 대화보다는 수 배에 달하게 획기적인 음성 데이터 전송률 감소 성능을 보이는 것을 알 수 있다.

IV. Result

다음은 본 연구의 구현 결과이다. 에이전트 학습 후, 텐서 보드 (Tensor Board)를 통해 결과를 시각화하였다.

Fig. 6. UI that intuitively shows that the user is speaking on the screen when speaking in voice conversation in response to text conversation

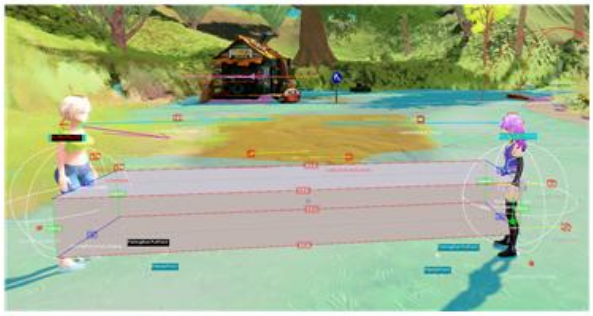

Fig. 7. Create a cube in the editor and measure the length of the cube

Fig. 8. After confirming that the distance between users operating on the system matches, proceed with distance-related measurements

Fig 6, 7, 8은 제안 시스템에서 개발한 결과를 보여주고 있다. 서버와 연동되는 클라이언트의 경우 다음과 같은 기능을 포함한 개발을 진행하였다.

① 로그인 매니지먼트 - 언제 사용자를 음성 대화 서버에 연결하고 대화를 가능하게 하는지 결정하는 로직 포함

② 채널 매니지먼트 - 어떤 채널을 사용하고 언제 참여하고 떠나는지를 대해 결정하는 로직 포함

③ 에러 핸들링 - 예기치 못한 통신장애와 같은 상황에서의 핸들링 방식 포함

일반 PC의 경우 기본 미디어 및 음향을 듣기 위한 오디오 디바이스 선택이 있지만, 본 환경에서는 음성 대화 환경 등을 고려한 헤드셋 등을 사용할 때를 대비한 커뮤니케이션 모드용 오디오 디바이스 선택이 따로 준비되어 있다. 사용자의 옵션 창에 일반 오디오용 디바이스 선택 및 음성 대화용 디바이스 선택을 따로 구분할 필요가 있는 만큼 커뮤니케이션 용 디바이스 선택 옵션 창을 따로 구성하여 옵션을 설정하였으며, 일반 오디오 디바이스와 음성 대화용 오디오를 중복해서 사용할 경우에 대한 시스템 설정 또한 고려한 구성을 진행하였다.

일반 모바일 기기는 자체적으로 음성 입력 도구와 출력 도구가 갖춰져 있지만, 무선 혹은 유선으로 헤드폰을 사용하는 경우 기기 선정의 우선순위에 따라 음성 출력 등에 여러 가지 변수가 발생할 수 있다. 이러한 환경에서는 가장 합리적이라고 생각되는 순서를 따라, 유용한 기기 선별 및 지정을 통한 기능 구현을 진행하였으며 3D 오디오가 기본 동작하는 환경을 고려하여 모든 음원은 모노(Mono)라는 상황으로 전제해 진행하였다.

V. Conclusions

본 연구는 근접 음성 대화 시스템 개발을 통해 대형 플랫폼에 의존하지 않고도 경제적으로 음성 대화 기능을 자체 구축할 방안을 모색하는 데 중점을 두었다. 이를 통해 소규모 회사나 개발자들이 과도한 비용 부담 없이 자체적인 음성 대화 기능을 구현할 가능성을 탐색한다. 근접 음성 대화 기능은 사용자들이 게임 내에서 물리적 거리에 따라 음성 소통이 가능하게 함으로써, 현실감 있는 상호작용을 제공하는 새로운 방식을 제시한다. 본 연구를 통해 개발된 근접 음성 대화 시스템에 대해 메타버스 콘텐츠, 잠입 대전 FPS 게임, 가상공간 공연, 가상 오피스의 리모트 회의, 그리고 지도 내 사용자 간의 약속 및 상대 찾기 등 다양한 활용 가능성을 찾을 수 있었다. 이는 음성 대화 기능의 개발과 서비스를 더욱 경제적으로 수행하려는 기업이나 개발자에게 실질적인 가이드라인을 제공하며, 새로운 음성 대화 기반 서비스의 개발을 촉진할 것으로 기대할 수 있다.

ACKNOWLEDGEMENT

본 연구는 중소벤처기업부와 중소기업기술정보진흥원의 “지역특화산업육성+(R&D, S3365329)”사업의 지원을 받아 수행된 연구결과임.

REFERENCES

[1] Valin, J.M., and T. Terriberry. "Opus Codec." Internet Engineering Task Force, RFC 6716, September 2012.

[2] Johnston, J.D. "Transform Coding of Audio Signals Using Perceptual Noise Criteria." IEEE Journal on Selected Areas in Communications 6, no. 2 (1988): 314-323.

[3] "Among Us: Enhancing Social Play with Proximity Voice Chat." Game Developer Conference, 2021.

[4] ITU-T Recommendation G.711. "Pulse code modulation (PCM) of voice frequencies." International Telecommunication Union, 1988.

[5] Perkins, C., O. Hodson, and V. Hardman. "A Survey of Packet Loss Recovery Techniques for Streaming Audio." IEEE Network 12, no. 5 (1998): 40-48.

[6] Bormann, C., and Z. Shelby. "Block-Wise Transfers in the Constrained Application Protocol (CoAP)." Internet Engineering Task Force, RFC 7959, August 2016.

[7] Boutremans, C., and J.-Y. Le Boudec. "Adaptive Joint Playout Buffer and FEC Adjustment for Internet Telephony." In INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, 3: 652-662.

[8] Salsano, S., L. Veltri, and D. Papalilo. "SIP: Session Initiation Protocol." RFC 3261, Internet Engineering Task Force, June 2002.

[9] WebRTC project. (2021). "WebRTC 1.0: Real-time Communication Between Browsers." W3C.

[10] Among Us. (2020). "Among Us: A Study on Proximity Voice Chat Impact on Online Games." InnerSloth.

[11] H.-N. Lee, D. Lee, and Y. I. Eom, “Analysis of Technical Trend for Scientific Networks,” Proceedings of the Korea Information Processing Society Conference, pp. 100-101, May 2013.

[12] OPUS, “https://www.opus-codec.org/”

Authors

Chief Executive Officer, Braves Co., ltd., Anyang, Korea. Jae-Woo Chang has a rich background in game content publication, media planning, and development. In 2000, He published a monthly game content magazine within KBS's Media division. His experience extends to media planning and management for a cable TV station, along with developing and managing several online gaming projects. Chang has also worked internationally, holding positions at Gamania in Taiwan and Gungho Online in Japan, where he collaborated with overseas developers in various capacities. Currently, he is the founder and CEO of Braves, a content development company based in Anyang.

Jin-Woong Kim received the B.S. degree in Computer Engineering, Jeju National University, Jeju, Republic of Korea, in 2023, where he is currently pursuing the master's degree in computer science. He interested in Artificial Intelligence Computing for Computer graphics, and Affordance of Graphic media such as game contents.

Soo Kyun Kim received Ph.D. in Computer Science & Engineering Department of Korea University, Seoul, Korea, in 2006. He joined the Telecommunication R&D Center at Samsung Electronics Co., Ltd., in 2006 and 2008. He is now a professor at the Department of Computer Engineering at Jeju National University, Korea. Dr. Kim has published many research papers in international journals and conferences. His research interests include multimedia, pattern recognition, image processing, mobile graphics, geometric modeling, and interactive computer graphics. He is a member of ACM, IEEE, IEEE CS, KACE, KMMS, KKITS, and KIIT.