

119 STT 처리능력 및 안정성 검증을 위한 성능부하시험장치 개발에 관한 연구

김윤수, 정병호*, 임정현, 홍윤기

(주)위니텍

kys7532@winitech.com, *usire@winitech.com

A Study on the Development of a Performance Load Testing Device for Evaluating the Processing Capability and Stability of 119 STT

Kim Yoonsu, Chung Byungho*, Lim Junghyun, Hong Yoonki,

Winitech Co., Ltd.

요 약

Abstract

대규모 재난이 발생하면 119 상황실로 대량의 신고가 동시에 들어오게 된다. 재난 현장의 소음과 신고접수 상황실의 혼란, 접수자의 숙련도 부족이나 누적된 피로에 의해 전화 신고 내용을 부정확하게 인지하게 되는 경우가 발생한다. 이러한 상황을 보완 하고자 신고전화에 대한 실시간 음성인식(Speech to text, STT) 기술을 도입하였으나, 대량 트래픽이 발생하는 고부하 환경에서는 처리 지연과 인식 오류가 발생할 수 있다. 이러한 문제는 신고 접수 지연과 상황 판단 오류로 이어져 재난 대응의 신속성과 정확성을 저하시킬 우려가 있다. 본 연구에서는 이를 사전에 검증하기 위한 오픈소스 Asterisk IP-PBX 플랫폼을 기반으로 한 저비용·정량형 STT 성능부하시험 장치를 제안한다. 제안된 시스템은 엡지 컴퓨팅 모듈과 다수의 시뮬레이션 노드를 연동하여 다중 콜 환경을 인위적으로 재현하고, STT 시스템의 처리 지연·오류율·안정성 등 주요 성능 지표를 정량적으로 검증한다. 이를 통해 고가의 상용 장비 없이도 재난 대응 기관이 독립적으로 운용 가능한 시험 환경을 구축할 수 있으며, 향후 AI 콜봇 및 자동 음성응답(IVR) 등 음성 기반 서비스의 성능 검증에도 확장 적용이 가능하다.

I. 서 론

최근 재난·안전 분야에서는 신속하고 정확한 음성 기반 신고 및 지휘 체계의 지능화가 중요한 기술 과제로 부상하고 있다. 특히 대규모 재난이 발생하면 119 상황실과 같은 긴급 대응 기관에는 동시에 수십 건 이상의 신고가 접수되며, 이로 인해 음성 데이터의 실시간 처리량이 급격히 증가한다. 이 과정에서 신고자의 발화가 격앙되거나 현장의 소음이 혼입되는 등 음성 품질이 저하되는 상황이 빈번히 발생하며, 이러한 환경에서는 수신자의 피로도와 업무 부담이 증가하여 정보 누락이나 오인식이 발생할 가능성이 커진다.

이러한 문제를 보완하기 위해 음성인식(Speech-to-Text, 이하 STT) 기술이 도입되어 신고자의 발화 내용을 실시간으로 텍스트로 변환하고, 이를 기반으로 빠른 상황 판단과 의사결정을 지원하는 체계가 확산되고 있다. 그러나 STT 기술은 주로 인식 정확도(Word Error Rate, Accuracy 등)를 중심으로 평가되어 왔으며, 실제 운영 환경에서 발생하는 대규모 동시 통화·고부하 트래픽 상황에서의 처리 속도, 안정성, 병목 현상 등을 검증할 수 있는 체계는 부재하다.

따라서 STT 기술이 재난 대응 현장에서 안정적으로 운용되기 위해서는 인식 품질뿐 아니라 대량 트래픽 처리 능력과 시스템 안정성을 객관적으로 평가할 수 있는 별도의 검증 체계가 필요하다. 본 연구에서는 이러한 기술적 공백을 해결하기 위해, 실제 재난 상황의 신고 폭주 환경을 모사할 수 있는 STT 성능부하시험 장치를 개발하고 그 구조와 검증 방법을 제안한다.

II. 본론

2.1 관련 기술 및 시스템 구성

본 연구는 재난 대응 기관의 음성 기반 신고 체계에서 발생하는 다중 신고 및 트래픽 폭주 상황을 정량적으로 검증하기 위해, STT 처리 과정과 연계된 음성 데이터 전송 인프라의 부하·처리 안정성 평가 장치를 제안한다. 제안된 시스템은 오픈소스 IP-PBX(Asterisk) 플랫폼을 기반으로 통화 제어, 음성 데이터 수집, 부하시험 제어 및 성능 모니터링 기능을 통합하였다. 전체 구조는 중앙 제어를 담당하는 고성능 엡지 컴퓨팅 모듈과, 실제 다중 부하 환경을 재현하는 경량형 발신 시뮬레이션 노드로 구성된다. 엡지 모듈은 SIP(Session Initiation Protocol) 기반 통화 세션을 관리하고, 음성 데이터를 표준화된 입력 형식으로 가공한 뒤 부하시험 시나리오에 따라 전송한다.[2],[3]

시뮬레이션 노드는 다수의 발신 요청을 자동 수행함으로써, 실제 119 상황실에서의 신고 폭주 환경과 유사한 네트워크 부하 조건을 생성한다.

이러한 구조를 통해 본 장치는 STT 시스템의 품질 자체를 평가하는 장비가 아니라, STT와 연계되는 통신 인프라의 처리 안정성, 지연, 세션 유실율, 패킷 손실률 등 성능 지표를 정량적으로 분석할 수 있다. 또한, 상용 고가 장비 없이도 실제 운용 환경을 모사할 수 있는 개방형 시험 환경을 구축할 수 있어, 향후 공공기관이나 지자체에서도 독립적으로 성능 검증을 수행할 수 있는 기반 기술로 활용 가능하다.

2.1.1 사설교환기(PBX) 시스템 설계

기존 상용 PBX 장비는 고비용·폐쇄형 구조로 대규모 동시 통화를 처리하기 어렵다. 본 연구에서는 오픈소스 Asterisk 플랫폼을 활용하여 SIP 등록, IVR(자동응답), 통화 제어, 음성 데이터 수집 기능을 통합한 Compact

PBX 시스템을 설계하였다.[1][2] 각 통화 세션에서 발생하는 RTP 스트림은 실시간으로 추출되어 PCM/WAV 형식으로 변환·저장되며, 이는 STT 처리 시스템의 안정성 시험을 위한 입력 데이터로 활용된다.[1]

PBX 내부에는 MQTT(Message Queuing Telemetry Transport) Broker가 내장되어 있으며, 콜 세션의 상태, 지연 시간, 패킷 손실 등 운용 데이터를 실시간으로 송·수신한다.[3] 수집된 데이터는 내부 데이터베이스에 자동 저장되어, 콜 수립률, CPU 점유율, 세션 유지율 등 시스템 처리 성능 분석에 활용된다. 옛지 모듈은 전체 통화 흐름의 제어와 데이터 교환의 중심 역할을 수행하며, 시뮬레이션 노드는 SIP 트래픽을 지속적으로 발생시켜 실제 신고 폭주 환경을 재현한다. 이 같은 구조는 재난 대응 환경에서의 음성처리 시스템의 부하 한계와 안정성을 정량적으로 평가할 수 있다.

2.1.2 부하시험 시스템의 동작 원리

성능부하시험 시스템은 오픈소스 IP-PBX를 기반으로 통화 처리, 음성 수집, 부하 제어, 성능 측정 기능을 통합한 소프트웨어 기반 검증 구조를 가진다. SIP 세션 제어와 IVR 기능을 통해 다중 발신·응답 시나리오를 자동 수행하며, 오디오 프레임 단위로 수집된 음성 데이터는 PCM/WAV 포맷으로 변환되어 시스템 처리 시험의 입력으로 사용된다.[2] 웹 기반 모니터링 화면을 통해 콜 수립률, CPU 점유율, 시스템 부하 상태를 실시간 확인할 수 있고, SIP 프레임워크를 활용하여 동시콜·DTMF 응답·콜 폭주 상황을 자동 시뮬레이션한다. 시험 결과는 로그로 집계되어 CPU 부하율, 처리 지연, 데이터 손실률 등 주요 지표를 분석할 수 있으며, 코덱 변환이나 버퍼 오버플로우 등으로 인한 음성 품질 저하도 함께 관찰할 수 있다. 이 시스템은 특정 STT 엔진이나 음성인식 알고리즘에 종속되지 않고, 표준 프로토콜 기반의 데이터 교환 구조를 통해 다양한 STT 기반 음성처리 환경에서 활용 가능하다.

2.2 시스템 구성 및 부하시험 시나리오

2.2.1 시스템 구현 및 데이터 흐름

전체 시스템은 중앙 제어 노드(옛지 컴퓨팅 모듈)와 부하시험 단말(시뮬레이션 노드)로 구성된다. 중앙 노드는 SIP 통화 관리, 음성 수집, 메시지 브로커 통신을 수행하고, 여러 개의 시뮬레이션 노드는 SIP 등록·발신·종료를 반복하며 초당 수십~수백 건의 트래픽을 발생시킨다.[1][2] 이 분산형 구조를 통해 실제 119 신고 폭주 상황을 재현할 수 있으며, 웹 대시보드를 통해 CPU 점유율, 콜 수립률, 패킷 손실률 등을 실시간으로 모니터링한다.

2.2.2 부하시험 시나리오 설계

본 연구에서 제안하는 부하시험은 Load, Stress, Spike, Endurance의 네 가지 시나리오로 구성되며, 각 시나리오는 STT 기반 음성처리 시스템의 트래픽 대응 능력과 운영 안정성을 정량적으로 검증하는 것을 목표로 한다.

이 네 가지 시나리오를 조합하여 수행함으로써, 시스템은 동시 세션 처리 성능·지연 특성·패킷 손실률·콜 수립 안정성 등 통신 품질 중심의 핵심 지표를 다각도로 검증할 수 있다. 이를 통해 제안된 시스템이 실제 119 상황실과 같은 대규모 동시 신고 환경에서 안정적으로 동작하는지를 실증적으로 평가할 수 있다. 특히 각 시험 결과는 CPU 부하율, 세션 유지율, 응답 지연 등의 정량 데이터를 통해 시각화되며, STT 시스템의 인식 품질이 아닌 처리 안정성을 중심으로 지표를 설정했다는 점에서 차별성을 가진다.

III. 결론

본 연구는 재난·안전 분야에서 운용되는 119 음성인식(Speech-to-Text; STT) 기반 신고 체계의 처리능력과 운영 안정성을 정량적으로 검증하기 위한 오픈소스 기반 성능부하시험 시스템의 구조와 구현 방안을 제시하였다. 제안된 시스템은 IP-PBX 플랫폼을 기반으로 SIP 통화 제어, 음성 데이터 수집, 메시지 브로커(MQTT) 기반 통신 기능을 통합적으로 수행하며, 고성능 옛지 컴퓨팅 모듈과 다수의 경량형 시뮬레이션 노드를 활용하여 대규모 동시 통화 및 트래픽 부하 환경을 실감형으로 재현하였다. 이를 통해 고가의 상용 장비 없이도 STT 시스템이 실제 재난 상황에서 지연 없이 안정적으로 처리 가능한지를 정량적으로 검증할 수 있는 개방형 시험 플랫폼을 구현하였다.

이번 연구는 기존의 음성인식 정확도 중심 평가를 넘어, 재난 대응 환경에서의 시스템 처리 안정성 검증 체계를 정립했다는 데 있다. 특히, 분산형 구조를 통해 다중 콜 트래픽과 부하 변동을 유연하게 모사할 수 있어, 공공 안전통신 인프라의 안정성 확보와 시스템 신뢰성 향상에 기여할 수 있다.

향후에는 실제 119 상황실의 운용 데이터를 기반으로 콜 패턴, 네트워크 부하, 코덱 변환 손실 등 세부 항목별 성능 검증 기능을 고도화할 예정이다. 또한 STT 시스템의 상태 정보를 실시간으로 수집하고, 이를 제어장치가 즉시 받아 분석·조정하는 자동 피드백 구조를 MQTT 통신 기반으로 확장할 계획이다. 이를 통해 처리 지연, 세션 유지율, 패킷 손실 등 주요 성능 지표를 즉시 확인하고 시스템 안정성을 향상시킬 수 있다.

궁극적으로 본 연구는 공공 및 민간 부문에서 운용되는 음성 기반 AI 서비스의 안정성 검증 및 품질 평가 인프라로 확장 가능한 기술적 기반을 제공한다는 점에서 의의가 있다.

ACKNOWLEDGMENT

This research was supported by the Regional Innovation System & Education(RISE) program through the Daegu RISE Center, funded by the Ministry of Education(MOE) and the Daegu Metropolitan City, Republic of Korea [2025-RISE-03-001]

참 고 문 헌

- [1] Costa, L. R., Nunes, L. S. N., Bordim, J. L., & Nakano, K. (2015, May). Asterisk PBX capacity evaluation. In 2015 IEEE International Parallel and Distributed Processing Symposium Workshop (pp. 519-524). IEEE.
- [2] Hoang, H. S., Tran, A. K., Doan, T. P., Tran, H. K., Dang, N. M. D., & Nguyen, H. N. (2023). Design and implementation of a VoIP PBX integrated Vietnamese virtual assistant: a case study. Journal of Information and Telecommunication, 7(2), 201-226.
- [3] Shahri, E., Pedreiras, P., & Almeida, L. (2022). Extending MQTT with real-time communication services based on SDN. Sensors, 22(9), 3162.

Table 1. Performance Load Testing Scenarios

Category	Development Details
① Load Testing	- Generate continuous traffic exceeding the expected usage to measure performance metrics such as call establishment rate, DTMF response time, and CPU utilization.
② Stress Testing	- Gradually apply call requests beyond the maximum capacity to analyze system bottlenecks and failure points under overload conditions.
③ Spike Testing	- Rapidly increase traffic within a short period to evaluate the system's instantaneous responsiveness and recovery capability.
④Endurance Testing	- Maintain high-load conditions over an extended period to monitor memory leaks, call loss rate, and system malfunction occurrences.