

English

1. Summary of Accomplishments

- Migrated Lex V2 (**ko_KR**) bot & flow to **ap-northeast-2 (Seoul)**; fixed locale/region mismatch and stabilized calls.
- Implemented **progressive streaming playback** (μ -law 8 kHz WAV) with **first-part quick start** and enforced **play-first** → **then advance index** ordering.
- Switched contact flow to **presigned S3 URL playback** and **stopped per-part Prompt creation**; freed quota using a **bulk deletion script** (Description contains “Dynamic TTS”).
- Added **demo-safe fallback** in Lambda (speaks **FallbackText** if TTS/S3 is unavailable).
- Updated documentation (**README.md**, **README_ko.md**) and prepared a **Twilio pilot kit** (Flask TwiML + Media Streams WS) for optional trials.

2. Issues & Risks

- **Prompt quota** can be exhausted if per-part prompts return; mitigation is to keep **URL playback** and run nightly prompt GC if prompts are reintroduced.
- **True “barge-in + process same utterance”** is constrained in Connect+Lex; awaiting **Chime SDK Voice (SMA)** enablement as the long-term fix (**StartBotConversation** → **CallAndBridge** to Connect).
- **Latency spikes** on TTS/S3 may delay first audio; mitigated by short quick-poll, retry, and **FallbackText** path.

3. Next Steps

- When AWS enables **SMA** in Seoul: implement **StartBotConversation(ko_KR)** then **CallAndBridge** to Connect; validate mid-prompt barge-in carrying the same utterance.
 - (If prompts are used again) adopt **one-prompt-per-contact** with **UpdatePrompt + delete at hangup**; avoid per-part prompts permanently.
 - Add telemetry: E2E turn latency, time-to-part0, error rates; optional Twilio Gather/Streams demo behind a reverse proxy when needed.
-

한국어

1. 주요 성과

- Lex V2(**ko_KR**) 봇과 플로우를 서울(ap-northeast-2)로 정리하여 지역/로케일 불일치 문제를 해소하고 안정화.
- μ-law 8kHz 스트리밍 재생과 첫 파트 빠른 시작 구현, “재생 후 인덱스 진행” 순서를 강제하여 끊김/누락 방지.
- 컨택트 플로우를 프리사인 S3 URL 재생으로 전환하고 퍼-파트 **Prompt** 생성 중단; 설명에 “Dynamic TTS”가 포함된 프롬프트를 대량 삭제 스크립트로 정리.
- Lambda에 데모 안전(**FallbackText**) 경로 추가(TTS/S3 불가 시 정중한 안내 후 종료).
- 문서(**README.md**, **README_ko.md**) 업데이트 및 **Twilio** 파일럿 키트(Flask TwiML + Media Streams WS) 준비.

2. 문제/리스크

- 프롬프트 쿼터 재발 위험 → **URL** 재생 유지 및 (재도입 시) 야간 정리로 예방.
- **“바지-인 + 동일 발화 처리”**는 Connect+Lex 단독 제약 존재 → **Chime SMA** 활성화 필요 (장기 해법).
- **TTS/S3** 자연으로 첫 재생 지원 가능 → 짧은 폴링/재시도 및 **FallbackText**로 완화.

3. 다음 단계

- SMA 활성화 후 **StartBotConversation(ko_KR)** → 완료 시 **CallAndBridge** → **Connect**로 연결, 중간 프롬프트 중 발화도 동일 스트림으로 처리 검증.
- (프롬프트 재도입 시) 세션당 1 개 **Prompt**만 생성 후 **UpdatePrompt** 사용, 통화 종료 시 삭제.
- 지표 강화: 턴당 자연, part0 도착시간, 실패율; 필요 시 리버스 프록시 뒤에서 Twilio Gather/Streams 데모 병행.