

전체 흐름 (General Flow)

[고객]

ㄷ 인바운드 또는 아웃바운드 전화 수신

[AWS Connect Contact Flow]

- ▶ [1. PlayPrompt: 홍보 오디오 재생 (Amazon Polly 또는 S3 사전 녹음 파일)]
 - └ 인터럽트 허용: 고객이 말하면 재생 중단
- ▶ [2. 고객 발화 수신 (STT)]
 - └ Amazon Lex 또는 Amazon Transcribe 사용
- ▶ [3. Lambda 호출 → AI 응답 생성]
 - └ STT 결과(고객 발화)를 전달
 - └ GPT 또는 사내 챗봇 호출
 - └ 응답 + 종료 여부(`is_end`: true/false) 반환
- ▶ [4. 음성 응답 생성]
 - └ 초기: Amazon Polly 사용
 - └ 이후: FishSpeech (자체 클론 보이스 API)로 대체
- ▶ [5. 응답 오디오 재생 (S3 URL을 통한 PlayPrompt)]
- ▶ [6. 종료 조건 판단]
 - └ `is_end == true` → 종료 멘트 재생
 - └ `is_end == false` → 다시 [2]로 반복
- ▶ [7. 상담원 연결 (필요 시)]
 - └ Amazon Connect 상담사 또는 PSTN으로 연결

💾 모든 고객 발화 및 AI 응답은 DynamoDB 또는 RDS에 고객 ID 및 타임스탬프와 함께 저장됩니다.

✉ 종료 시점에 SMS 발송 가능 (Amazon SNS 또는 자체 API 사용)

✅ 구현 완료 항목 (Finished Implementations)

- 통화 연결 시 기본 홍보 메시지 자동 재생
- 자연스러운 음성 합성(TTS) 구현 완료

🔧 향후 작업 항목 (To-Do List)

TTS (FishSpeech 관련)

- FishSpeech 기반 다양한 음성 프롬프트 생성 (유사 화자 참조 활용)

- FishSpeech 추론 속도 개선
- 생성된 음성을 S3에 자동 저장하고, Contact Flow에서 사용할 수 있도록 자동 등록 스크립트 작성





AWS Connect 관련 (향후 점진적 제거 예정)

- 고객 발화 STT 처리 → Lambda 통합
 - 고객 발화 시 Contact Flow 동작 구성:
 1. 홍보 재생 중단 (인터럽트)
 2. 고객 응답에 따른 음성 프롬프트 재생 (SMS 외 케이스). 추후 AI Agent로 로직 대체 예정
 3. 특정 조건 충족 시 고객에게 SMS 발송
 4. 고객 발화 지속 시, 시스템이 의도를 판단하여 상담원으로 연결 (에스컬레이션)
 - 전체 음성과 텍스트 흐름 저장 (고객 ID + 타임스탬프 기반 식별자 포함)
-

AI 에이전트 관련 고려사항

- 높은 수준의 STT 정확도 확보
 - AI 에이전트가 고객 응답에 따라 적절한 조치를 자동 판단 및 수행:
 - 음성 응답 (TTS)
 - 문자 발송 (SMS)
 - 상담원 연결 여부 판단
 - 수동 응답 또는 fallback 로직은 사용하지 않으며, 모든 응답 처리는 AI 기반으로 수행됨
-

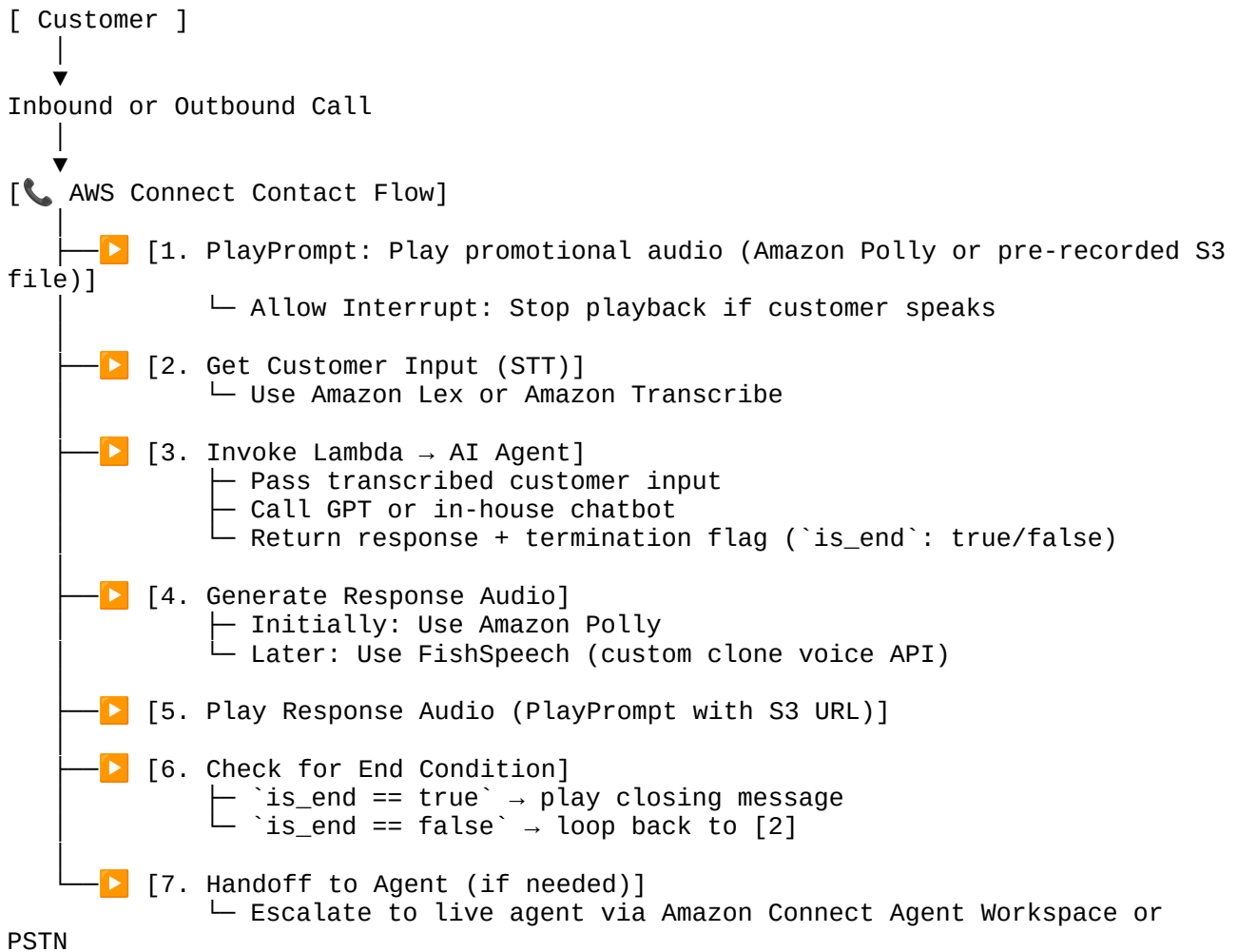
예상 추진 일정 (Tentative Next Steps)

-  [즉시] 고객 발화 시 실시간 반응 가능하도록 구현 (재생 중단 및 STT 처리 연동)
 -  [단기~중기] STT 정확도 개선 및 흐름 내 자연스러운 텍스트 변환 처리
 -  [중기~장기] 전체 흐름 검증 (SMS, 다양한 음성 프롬프트, 상담사 연결 포함)
 -  [지속적] 전체 자동화 흐름 스크립트화 및 문서화. 오류 최소화 및 반복 작업 제거 목표
-

시스템 단계별 전환 계획 (System Migration Plan)

- 1 단계: 초기에는 Amazon Polly (TTS), Amazon Lex (챗봇) 사용
- 2 단계: Polly → FishSpeech(자체 음성 합성 엔진)로 대체
- 3 단계: Lex → 사내 개발 GPT 기반 챗봇으로 전환하여 커스터마이징 및 독립성 확보

General Flow



💾 All customer inputs and AI responses are logged to DynamoDB or RDS with a unique identifier (customer ID + timestamp)

✉️ Final messages can be sent via SMS (Amazon SNS or custom API)

✅ Finished Implementations

- Promotional message playback triggered upon call connection
 - Natural-sounding speech generation via TTS completed
-

🔧 To-Do List

TTS (FishSpeech)

- Generate diverse voice prompts using FishSpeech (using reference voice)
- Improve inference speed for FishSpeech generation

- Script automation to save generated prompts to S3 and register them for use in Amazon Contact Flow

AWS Connect (to be minimized)

- Integrate customer speech processing (STT) into Lambda
 - Contact Flow behavior when customer speaks:
 1. Interrupt and pause promo playback
 2. Play appropriate voice prompt based on customer response (non-SMS cases); later replace logic with AI Agent
 3. Trigger SMS to customer when conditions are met
 4. If conversation continues, determine if escalation to live agent is required
 - Log complete interaction flow (voice + text) with traceable identifiers
-







AI Agent Considerations

- Ensure high STT transcription accuracy
 - The AI Agent must autonomously determine the appropriate action:
 - TTS response
 - SMS response
 - Escalation to live agent if necessary
 - No manual fallback logic is used; all response handling is AI-driven
-



Tentative Next Steps

-  [Immediate] Support real-time customer speech during promo; auto-interrupt and process via STT
 -  [Short-to-Mid Term] Complete STT integration into call flow for accurate transcription
 -  [Mid-to-Long Term] Validate full end-to-end flow: prompt switching, SMS triggers, escalation
 -  [Ongoing] Provide technical documentation and automation scripts to streamline operations and reduce manual steps
-



System Migration Plan

- **Stage 1:** Initial deployment using Amazon Polly (TTS) and Amazon Lex (chatbot)
- **Stage 2:** Replace Polly with FishSpeech, our internal voice synthesis engine
- **Stage 3:** Replace Lex with our own GPT-based chatbot for full control and scalability