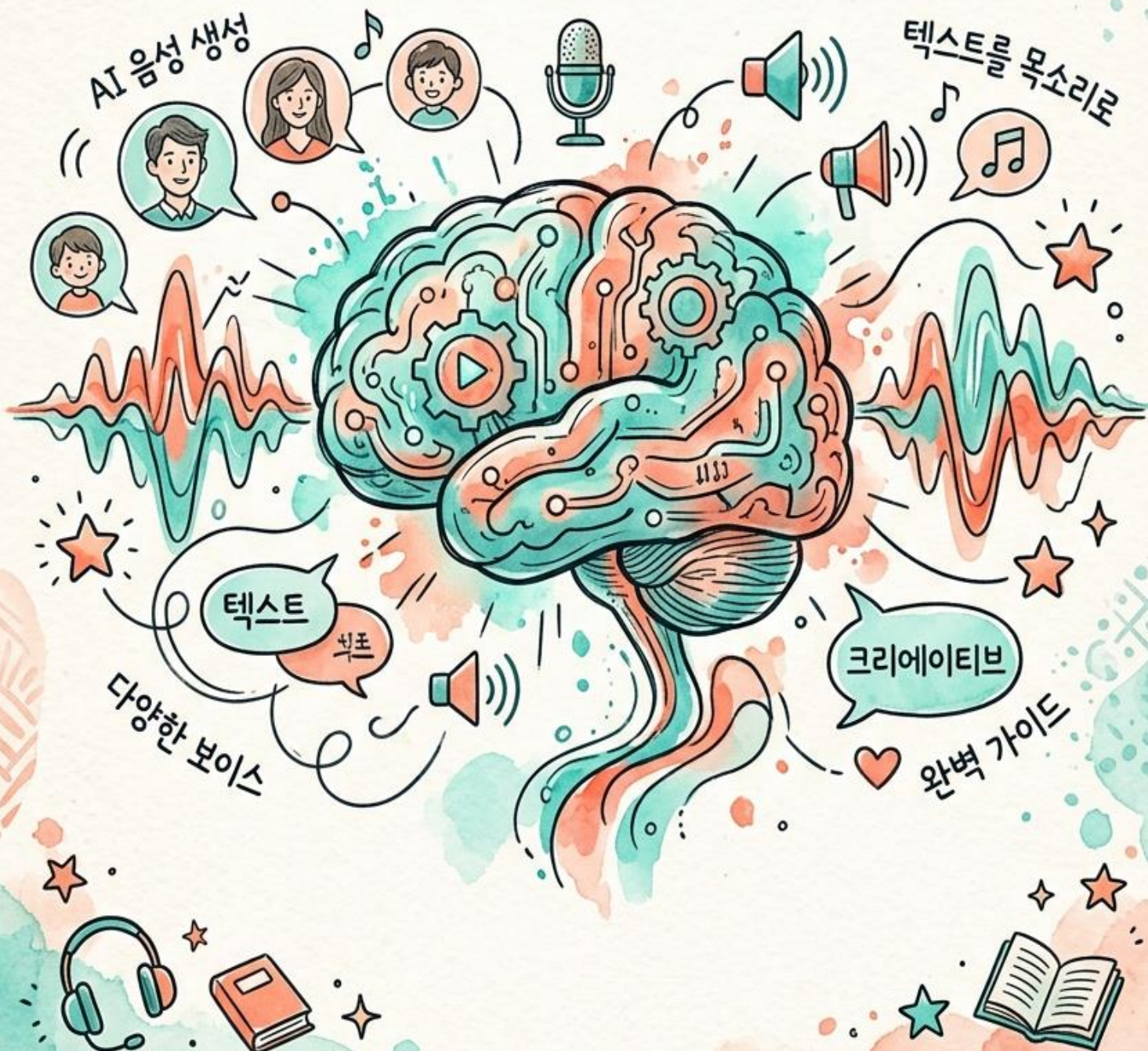


일레븐랩스 100% 활용법



※ 이 책은 AI의 도움을 받아 작성되었습니다. 일부 오류가 있는 부분이 있으니 참고용으로 아이디어 및 개요 관점으로 봐주시 바랍니다. 또한, 시점에 따라 업데이트가 안된 부분이 있을 수 있으니 자세한 내용은 젠스파크 공식 문서를 확인하여 주시기 바랍니다.

[더 많은 정보]

네이버 프리미엄 콘텐츠 주소: <https://contents.premium.naver.com/vbcoding/vibecoding>

스레드 채널 주소: <https://www.threads.com/@vibe.coding.developer>

서문

이 책은 일레븐랩스(ElevenLabs)라는 AI 음성 기술 플랫폼의 모든 기능을 철저히 분석하고, 무료 플랜부터 유료 플랜까지 각 단계에서 최대한의 가치를 끌어내는 방법을 다룹니다.

2026년 현재, AI 음성 기술은 콘텐츠 제작, 교육, 비즈니스, 엔터테인먼트 등 거의 모든 영역에서 빠르게 자리를 잡아가고 있습니다. 유튜브 나레이션, 팟캐스트, 오디오북, 광고 보이스오버, 다국어 더빙, AI 음성 에이전트에 이르기까지, 사람의 목소리가 필요한 모든 곳에서 AI 음성이 활용되기 시작했습니다.

그 가운데 일레븐랩스는 "가장 자연스러운 AI 음성"이라는 평가와 함께 업계를 선도하고 있습니다. 텍스트를 음성으로 변환하는 TTS(Text-to-Speech)를 핵심으로 하면서, 음성 복제(Voice Cloning), 다국어 더빙(Dubbing), 음성 변환(Voice Changer), 음성 분리(Voice Isolator), 효과음 생성(Sound Effects), AI 음악 생성(Eleven Music), 음성 인식(Speech-to-Text), 대화형 AI 에이전트(Conversational AI), 보이스오버 스튜디오(Voiceover Studio) 등 음성과 오디오에 관한 거의 모든 기능을 하나의 플랫폼에서 제공합니다.

그런데 많은 사용자가 일레븐랩스의 잠재력을 충분히 활용하지 못하고 있습니다. 무료 플랜의 월 10,000자(Multilingual 모델 기준) 제한 안에서 어떻게 효율적으로 사용해야 할지 모르는 분, 음성 복제와 더빙 같은 고급 기능의 구체적인 활용법을 알지 못하는 분, TTS 결과물의 품질을 끌어올리는 텍스트 작성 기법을 모르는 분이 많습니다. 또한 유료 플랜으로 전환했다라도 크레딧을 전략적으로 관리하지 않으면 예상보다 빠르게 소진되는 경험을 하게 됩니다.

이 책은 바로 그런 분들을 위해 만들어졌습니다. 무료 플랜에서의 크레딧 절약 전략부터, 각 기능의 구체적인 사용법, TTS 품질을 극대화하는 텍스트 작성 기법, 음성 복제의 실전 노하우, 다국어 더빙 워크플로우, 대화형 AI 에이전트 구축법, 그리고 유료 플랜 선택 기준까지 일레븐랩스 활용에 필요한 모든 것을 다룹니다.

이 책에서 다루는 모든 내용은 2026년 3월 기준 일레븐랩스의 공식 문서(Documentation), 헬프센터, 공개된 사용법 가이드, 그리고 실제 사용 경험을 토대로 작성되었습니다. 특정 기업이나 서비스에 대한 부당한 비교나 비방은 포함하지 않았으며, 스크린샷 대신 텍스트 기반의 상세한 설명으로 구성하여 저작권 문제가 발생하지 않도록 했습니다.

목차

제1부. 일레븐랩스 기초 완전 이해

- 제1장. 일레븐랩스란 무엇인가
- 제2장. 요금제 구조와 크레딧 시스템 완전 해부
- 제3장. 계정 생성부터 초기 설정까지
- 제4장. 인터페이스와 대시보드 구조 파악하기

제2부. 핵심 기능별 완전 활용 전략

- 제5장. 텍스트 투 스피치(Text-to-Speech) — 핵심 중의 핵심
- 제6장. 음성 복제(Voice Cloning) — 나만의 AI 음성 만들기
- 제7장. 보이스오버 스튜디오(Voiceover Studio) — 장편 콘텐츠 제작의 중심
- 제8장. 다국어 더빙(Dubbing) — 언어의 벽을 넘는 기술
- 제9장. 음성 변환(Voice Changer) — 목소리를 자유자재로
- 제10장. 음성 분리(Voice Isolator) — 깨끗한 음성 추출
- 제11장. 효과음 생성(Sound Effects) — AI로 만드는 사운드
- 제12장. AI 음악 생성(Eleven Music) — 텍스트로 작곡하기
- 제13장. 음성 인식(Speech-to-Text) — AI 전사의 새로운 기준
- 제14장. 대화형 AI 에이전트(Conversational AI) — 음성 AI 비서 구축
- 제15장. 오디오 네이티브(Audio Native) — 웹 콘텐츠의 음성화
- 제16장. 보이스 라이브러리(Voice Library)와 아이코닉 마켓플레이스

제3부. 품질 극대화를 위한 고급 전략

- 제17장. TTS 품질을 극대화하는 텍스트 작성 기법
- 제18장. Eleven v3 모델의 오디오 태그와 감정 표현 마스터하기
- 제19장. 음성 설정 파라미터 최적화 — 안정성, 유사도, 스타일, 속도
- 제20장. 텍스트 정규화(Text Normalization) — 숫자, 약어, 기호 처리법
- 제21장. 발음 사전(Pronunciation Dictionary) 활용법

제4부. 크레딧 관리와 비용 최적화 전략

- 제22장. 크레딧 소모 패턴 분석과 최적 배분 전략
- 제23장. Flash 모델 vs Multilingual v2 — 상황별 현명한 선택
- 제24장. 무료 플랜 100% 활용법 — 월 10,000자의 전략적 사용
- 제25장. 초과 사용(Overage) 관리와 비용 예측

제5부. 실전 시나리오별 활용 가이드

- 제26장. 유튜버와 영상 크리에이터를 위한 활용 시나리오
- 제27장. 팟캐스터와 오디오 콘텐츠 제작자를 위한 시나리오
- 제28장. 교육자와 이러닝 콘텐츠 제작자를 위한 시나리오
- 제29장. 비즈니스와 마케팅 전문가를 위한 시나리오
- 제30장. 개발자와 기술 팀을 위한 API 활용 시나리오
- 제31장. 작가와 출판 관계자를 위한 오디오북 시나리오

제6부. 한계 인식과 현명한 대처

제32장. 일레븐랩스의 정직한 한계 분석

제33장. 유료 플랜 전환 시점 판단 기준과 플랜 선택 가이드

제34장. 상업적 이용과 저작권 — 알아야 할 것들

제7부. 부록

제35장. 자주 묻는 질문(FAQ) 50선

제36장. 기능별 크레딧 소모량 참고표

제37장. 용어 사전

제1부. 일레븐랩스 기초 완전 이해

1 일레븐랩스란 무엇인가

1.1 일레븐랩스의 탄생 배경과 핵심 철학

일레븐랩스(ElevenLabs)는 AI 기반 음성 기술에 특화된 플랫폼입니다. 이 회사의 핵심 철학은 “모든 콘텐츠를 모든 언어로, 모든 목소리로 전달할 수 있게 하는 것”입니다. 기존의 TTS(Text-to-Speech) 기술이 로봇 같은 어색한 발음에 머물러 있었다면, 일레븐랩스는 사람의 목소리와 거의 구별되지 않을 정도의 자연스러운 음성 합성 기술을 실현했습니다.

일레븐랩스가 다른 TTS 서비스와 구별되는 가장 핵심적인 차이는 감정의 표현력입니다. 단순히 텍스트를 읽어주는 수준을 넘어, 기쁨, 슬픔, 놀라움, 분노, 속삭임, 웃음 등 다양한 감정과 뉘앙스를 음성에 담아낼 수 있습니다. 이는 나레이션, 오디오북, 광고 등 감정 전달이 중요한 콘텐츠에서 결정적인 차이를 만들어냅니다.

1.2 일레븐랩스의 핵심 기능 전체 조망

일레븐랩스는 음성과 오디오에 관한 포괄적인 기능 세트를 제공합니다. 이 기능들을 크게 분류하면 다음과 같습니다.

음성 생성 영역에는 텍스트 투 스피치(TTS), 보이스오버 스튜디오, 음성 복제(인스턴트 클론 및 프로페셔널 클론), 보이스 디자인 등이 포함됩니다. TTS는 일레븐랩스의 가장 핵심적인 기능으로, 텍스트를 자연스러운 음성으로 변환합니다. 보이스오버 스튜디오는 긴 분량의 나레이션이나 오디오북 제작에 최적화된 타임라인 기반 편집 환경을 제공합니다.

음성 변환 및 처리 영역에는 음성 변환(Voice Changer), 음성 분리(Voice Isolator)가 있습니다. 음성 변환은 녹음된 음성의 특성을 다른 목소리로 바꾸주며, 음성 분리는 배경 소음에서 음성만 깨끗하게 추출하는 기능입니다.

다국어 처리 영역에는 자동 더빙(Dubbing)과 더빙 스튜디오(Dubbing Studio)가 있습니다. 29개 이상의 언어를 지원하며, 원본 음성의 톤과 감정을 유지하면서 다른 언어로 자동 번역 및 음성 합성을 수행합니다.

콘텐츠 생성 영역에는 효과음 생성(Sound Effects)과 AI 음악 생성(Eleven Music)이 포함됩니다. 텍스트 프롬프트로 효과음과 음악을 생성할 수 있어, 별도의 음향 라이브러리 없이도 콘텐츠에 필요한 오디오 요소를 만들 수 있습니다.

음성 인식 영역에는 음성 인식 및 전사(Speech-to-Text)가 있습니다. 일레븐랩스의 자체 음성 인식 모델인 Scribe를 통해 90개 이상의 언어에 대한 고정밀 전사를 제공합니다.

AI 에이전트 영역에는 대화형 AI(Conversational AI)가 있습니다. 실시간 음성 기반의 AI 에이전트를 구축할 수 있어, 고객 서비스, 교육, 인터랙티브 콘텐츠 등에 활용할 수 있습니다.

웹 연동 영역에는 오디오 네이티브(Audio Native)가 있습니다. 웹페이지에 임베드할 수 있는 오디오 플레이어를 제공하여, 블로그나 뉴스 기사 등의 텍스트 콘텐츠를 자동으로 음성화할 수 있습니다.

마켓플레이스 영역에는 보이스 라이브러리(Voice Library)와 아이코닉 마켓플레이스(Iconic Marketplace)가 있습니다. 커뮤니티가 공유한 수천 개의 음성을 무료로 사용할 수 있으며, 유명인의 공인된 AI 음성을 라이선스하여 사용할 수도 있습니다.

1.3 일레븐랩스의 플랫폼 접근성

일레븐랩스는 웹 브라우저 기반으로 작동하므로 별도의 소프트웨어 설치 없이 접속하여 사용할 수 있습니다. 데스크톱 컴퓨터, 노트북, 태블릿, 스마트폰 등 브라우저가 있는 모든 기기에서 접근이 가능합니다. 또한 모바일 앱(iOS 및 Android)도 제공되어 이동 중에도 TTS 생성, 음성 복제 등의 기능을 활용할 수 있습니다.

개발자를 위해서는 RESTful API와 다양한 언어의 SDK(소프트웨어 개발 키트)가 제공되므로, 자체 애플리케이션이나 서비스에 일레븐랩스의 음성 기술을 통합하는 것도 가능합니다. 이는 일레븐랩스를 단순한 도구를 넘어, 음성 기반 서비스의 인프라로 활용할 수 있게 해줍니다.

1.4 지원 언어와 글로벌 커버리지

일레븐랩스의 Multilingual v2 모델은 29개 이상의 언어를 지원하며, 한국어도 그 중 하나입니다. 영어, 한국어, 일본어, 중국어, 스페인어, 프랑스어, 독일어, 포르투갈어, 이탈리아어, 힌디어, 아랍어, 터키어, 폴란드어, 네덜란드어 등 주요 언어를 폭넓게 지원합니다.

한국어 TTS의 품질도 상당히 높은 수준에 도달해 있어, 한국어 콘텐츠 제작에도 충분히 활용 가능합니다. 다만, 영어에 비해서는 감정 표현의 범위나 자연스러움에서 약간의 차이가 있을 수 있으며, 이는 다국어 AI 모델의 일반적인 특성입니다.

2.1 크레딧이란 무엇인가

일레븐랩스에서 크레딧은 "문자 수(characters)"를 기본 단위로 하는 사용량 측정 체계입니다. TTS에서는 입력 텍스트의 문자 수로 크레딧이 소모되고, 더빙에서는 분(minutes) 단위로, 음성 인식에서는 시간(hours) 단위로, 효과음에서는 생성 횟수(generations) 단위로 각각 크레딧이 측정됩니다.

여기서 중요한 점은 "문자"란 글자, 공백, 구두점 등을 모두 포함한다는 것입니다. 예를 들어, "안녕하세요, 만나서 반갑습니다."라는 문장은 공백과 구두점을 포함하여 총 16자로 계산됩니다. 한국어의 경우 한 글자가 하나의 문자로 계산되며, 영어의 경우에도 하나의 알파벳이 하나의 문자로 계산됩니다.

대략적인 참고 기준으로, 영어 기준 약 1,000자가 대략 1분 분량의 음성에 해당합니다. 한국어의 경우 발화 속도와 문장 구조가 다르므로 정확히 같지는 않지만, 비슷한 수준으로 참고할 수 있습니다. 따라서 무료 플랜의 10,000자는 대략 10분 내외의 음성에 해당한다고 볼 수 있습니다.

2.2 요금제 전체 구조

일레븐랩스는 2026년 3월 현재 여러 단계의 요금제를 운영하고 있습니다. 각 요금제의 핵심 사항을 정리하면 다음과 같습니다.

무료(Free) 플랜은 월 0달러이며, Multilingual 모델 기준 월 10,000자(Flash 모델 기준 20,000자)가 제공됩니다. 음성 복제는 불가능하고, 상업적 사용이 금지되며(출처 표시 필요), 오디오 품질은 128kbps, 동시 요청은 2건까지 가능합니다. 커스텀 음성은 3개까지 생성할 수 있고, 프로젝트는 3개까지 만들 수 있습니다.

스타터(Starter) 플랜은 월 5달러(연간 결제 시 월 약 4.17달러)이며, Multilingual 기준 월 30,000자(Flash 기준 60,000자)가 제공됩니다. 인스턴트 음성 복제가 가능해지고, 상업적 사용 권한이 부여됩니다. 오디오 품질은 128kbps이며, 동시 요청은 3건까지 가능합니다.

크리에이터(Creator) 플랜은 월 11달러(첫 달 50% 할인 적용 시)이며, Multilingual 기준 월 100,000자(Flash 기준 200,000자)가 제공됩니다. 프로페셔널 음성 복제(PVC) 1개가 포함되고, 오디오 품질이 192kbps로 향상됩니다. 동시 요청은 5건이며, 초과 사용 시 1,000자당 0.30달러가 부과됩니다.

프로(Pro) 플랜은 월 99달러이며, Multilingual 기준 월 500,000자(Flash 기준 1,000,000자)가 제공됩니다. 프로페셔널 음성 복제 1개 포함, 192kbps 오디오 품질, 동시 10건 요청이 가능하며, 초과 사용 시 1,000자당 0.24달러가 부과됩니다.

스케일(Scale) 플랜은 월 330달러이며, Multilingual 기준 월 2,000,000자(Flash 기준 4,000,000자)가 제공됩니다. 3석(seats)이 포함되어 팀 사용이 가능하며, 동시 15건 요청, 초과 시 1,000자당 0.18달러입니다.

비즈니스(Business) 플랜은 월 1,320달러이며, Multilingual 기준 월 11,000,000자(Flash 기준 22,000,000자)가 제공됩니다. 프로페셔널 음성 복제 3개, 15석 이상 포함, SLA(서비스 수준 협약)가 제공되며, 초과 시 1,000자당 0.12달러입니다.

엔터프라이즈(Enterprise) 플랜은 맞춤 가격으로, 모든 사항이 협의를 통해 결정됩니다.

2.3 Flash 모델과 Multilingual 모델의 크레딧 차이

일레븐랩스의 TTS에는 두 가지 주요 모델이 존재합니다. Flash 모델과 Multilingual v2 모델입니다. 같은 요금제라도 사용하는 모델에 따라 사용 가능한 문자 수가 다릅니다.

Flash 모델은 속도에 최적화된 모델로, 지연 시간이 매우 짧아 실시간 대화형 AI에 적합합니다. Flash 모델을 사용하면 Multilingual 모델 대비 약 두 배의 문자를 사용할 수 있습니다. 예를 들어 무료 플랜에서 Multilingual은 10,000자이지만 Flash는 20,000자입니다.

Multilingual v2 모델은 품질에 최적화된 모델로, 29개 이상의 언어를 지원하며 감정 표현력과 억양의 자연스러움이 훨씬 뛰어납니다. 192kbps의 고품질 오디오를 출력할 수 있으며, 오디오북, 유튜브 나레이션, 팟캐스트 등 품질이 중요한 콘텐츠에 적합합니다.

이 차이를 이해하는 것은 크레딧 관리에 매우 중요합니다. 품질이 최우선인 최종 결과물에는 Multilingual 모델을, 초안 확인이나 테스트에는 Flash 모델을 사용하면 크레딧을 효율적으로 관리할 수 있습니다.

2.4 크레딧 갱신과 이월 규칙

크레딧은 결제 주기에 따라 월 단위로 갱신됩니다. 무료 플랜의 경우 가입일로부터 30일마다 갱신됩니다. 사용하지 않은 크레딧은 다음 달로 이월되지 않습니다. 이번 달에 5,000자만 사용하고 5,000자가 남았더라도, 다음 달에는 15,000자가 아닌 10,000자로 초기화됩니다.

유료 플랜에서도 마찬가지로, 기본 제공 크레딧은 이월되지 않습니다. 따라서 매달 주어진 크레딧을 효율적으로 사용하는 것이 중요합니다. 다만, 별도로 구매한 추가 크레딧이 있는 경우에는 기본 크레딧이 먼저 소진된 후 추가 크레딧이 사용되는 구조입니다.

2.5 한 번에 사용할 수 있는 크레딧 상한

남용을 방지하기 위해, 한 번의 생성 요청에서 사용할 수 있는 크레딧에 상한이 있습니다. 이 상한은 해당 플랜의 월간 크레딧의 두 배입니다. 예를 들어 스타터 플랜(30,000자)의 경우 한 번에 최대 60,000자까지 사용할 수 있고, 크리에이터 플랜(100,000자)의 경우 한 번에 최대 200,000자까지 사용할 수 있습니다.

또한 웹사이트에서 음성 합성(Speech Synthesis)을 사용할 때, 유료 플랜은 한 번에 최대 5,000자, 무료 플랜은 한 번에 최대 2,500자를 입력할 수 있습니다. 이보다 긴 텍스트를 변환해야 하는 경우에는 보이스오버 스튜디오나 API를 활용해야 합니다.

3 계정 생성부터 초기 설정까지

3.1 계정 생성 절차

일레븐랩스에 가입하는 과정은 매우 간단합니다. 웹 브라우저에서 일레븐랩스 공식 웹사이트(elevenlabs.io)에 접속합니다. 화면에서 가입 또는 시작하기 버튼을 클릭합니다. 구글 계정 연동, 깃허브(GitHub) 계정 연동, 이메일 주소를 통한 가입 등 여러 옵션이 제공됩니다.

무료 플랜 가입 시 신용카드 정보를 입력할 필요가 없습니다. 가입 즉시 월 10,000자 (Multilingual 기준)의 크레딧이 제공되며, TTS, 효과음 생성, 보이스 라이브러리 탐색 등 무료 기능을 바로 사용할 수 있습니다.

3.2 초기 설정 최적화

가입을 완료한 후 몇 가지 초기 설정을 확인하면 이후의 사용 경험이 향상됩니다.

인터페이스 언어를 확인합니다. 일레븐랩스의 인터페이스는 기본적으로 영어로 되어 있습니다. 현재 한국어 인터페이스는 제공되지 않을 수 있지만, AI 음성 생성 시 한국어 텍스트를 입력하고 한국어 음성을 출력받는 것은 문제없이 가능합니다.

기본 음성(Default Voice)을 설정합니다. TTS를 사용할 때마다 음성을 선택해야 하는데, 자주 사용하는 음성을 기본값으로 설정해두면 편리합니다. 보이스 라이브러리에서 다양한 음성을 미리 탐색하고, 자신의 용도에 적합한 음성을 "My Voices"에 저장해두는 것이 좋습니다.

크레딧 갱신일을 파악합니다. 가입일 기준으로 30일마다 크레딧이 갱신되므로, 자신의 갱신일을 알아두면 크레딧 사용 계획을 세우는 데 도움이 됩니다.

3.3 모바일 앱 설치

일레븐랩스는 iOS와 Android 모바일 앱을 제공합니다. 앱에서는 TTS, 인스턴트 음성 복제, 보이스 라이브러리 탐색 등 주요 기능을 사용할 수 있습니다. 이동 중에 빠르게 음성을 생성하거나 음성 샘플을 녹음하여 복제하는 등의 작업에 유용합니다.

다만, 프로페셔널 음성 복제(PVC), 더빙 스튜디오, 보이스오버 스튜디오 등 일부 고급 기능은 현재 웹 버전에서만 사용 가능할 수 있습니다. 모바일 앱의 기능은 지속적으로 확장되고 있으므로, 최신 업데이트를 확인하는 것이 좋습니다.

3.4 API 키 발급 (개발자용)

개발자이거나 자체 서비스에 일레븐랩스를 통합하려는 경우, API 키를 발급받아야 합니다. 대시보드의 설정(Settings) 메뉴에서 API Keys 항목으로 이동하면 API 키를 생성하고 관리할 수 있습니다. API를 통한 사용도 동일한 크레딧 풀에서 차감되므로, 크레딧 관리에 주의해야 합니다.

4 인터페이스와 대시보드 구조 파악하기

4.1 메인 대시보드의 구성

일레븐랩스에 로그인하면 메인 대시보드가 표시됩니다. 대시보드에서는 각 기능에 빠르게 접근할 수 있는 메뉴가 좌측 사이드바에 정리되어 있습니다. 주요 메뉴 항목으로는 텍스트 투 스피치(Text to Speech), 보이스오버 스튜디오(Voiceover Studio), 음성 복제(Voice Cloning), 더빙(Dubbing), 음성 변환(Voice Changer), 음성 분리(Voice Isolator), 효과음(Sound Effects), 음악(Music), 음성 인식(Speech to Text), 대화형 AI(Conversational AI), 보이스 라이브러리(Voice Library) 등이 있습니다.

4.2 텍스트 투 스피치 인터페이스

TTS 페이지에 접속하면 텍스트 입력 영역이 중앙에 위치합니다. 입력 영역 위나 옆에는 음성 선택 드롭다운, 모델 선택(Flash 또는 Multilingual 등), 음성 설정(안정성, 유사도, 스타일 등의 슬라이더), 속도 조절, 출력 형식 선택 등의 옵션이 배치되어 있습니다.

텍스트를 입력하고 "생성(Generate)" 버튼을 클릭하면, 입력된 텍스트에 해당하는 문자 수만큼 크레딧이 소모되며 음성이 생성됩니다. 생성된 음성은 브라우저 내에서 바로 재생하여 확인할 수 있고, 다운로드 버튼을 클릭하면 MP3 등의 형식으로 저장할 수 있습니다.

4.3 음성 선택과 관리

음성 선택 드롭다운을 클릭하면, 사용 가능한 음성 목록이 표시됩니다. 이 목록에는 일레븐랩스가 제공하는 기본 음성(Default Voices), 보이스 라이브러리에서 저장한 커뮤니티 음성, 그리고 사용자가 직접 복제하거나 디자인한 커스텀 음성이 포함됩니다.

“My Voices” 섹션에서는 자신이 사용하는 모든 음성을 관리할 수 있습니다. 자주 사용하는 음성은 즐겨찾기에 추가하여 빠르게 접근할 수 있도록 해두면 편리합니다.

4.4 이력 관리와 생성물 보관

생성한 음성 파일의 이력은 대시보드에서 확인할 수 있습니다. 이전에 생성한 음성을 다시 재생하거나 다운로드할 수 있어, 같은 작업을 반복할 필요가 줄어듭니다. 이는 크레딧 절약 면에서도 유용합니다. 이전에 만든 음성이 필요하다면, 다시 생성하지 않고 이력에서 다운로드하면 크레딧을 소모하지 않습니다.

4.5 크레딧 잔액 확인

크레딧 잔액은 대시보드의 프로필 영역이나 설정 메뉴에서 확인할 수 있습니다. 현재 남은 문자 수, 사용한 문자 수, 갱신일 등이 표시됩니다. 작업을 시작하기 전에 잔액을 확인하는 습관을 들이면, 작업 중 크레딧이 부족해지는 상황을 방지할 수 있습니다.

제2부. 핵심 기능별 완전 활용 전략

5 텍스트 투 스피치(Text-to-Speech) — 핵심 중의 핵심

5.1 TTS의 기본 원리

텍스트 투 스피치는 일레븐랩스의 가장 핵심적인 기능입니다. 텍스트를 입력하면 AI가 이를 분석하여 자연스러운 음성으로 변환합니다. 이 과정에서 AI는 텍스트의 문맥, 구두점, 문장 구조 등을 고려하여 적절한 억양, 리듬, 강세를 적용합니다.

일레븐랩스의 TTS가 특별한 이유는 단순한 “읽기”를 넘어 “연기”에 가까운 음성을 생성한다는 점입니다. 물음표가 있으면 울림 억양을, 느낌표가 있으면 강조를, 말줄임표가 있으면 자연스러운 멈춤을 적용합니다. 특히 최신 Eleven v3 모델에서는 오디오 태그를 통해 웃음, 속삭임, 한숨 등의 비언어적 표현까지 제어할 수 있습니다.

5.2 TTS 모델 선택 전략

일레븐랩스에서는 여러 TTS 모델을 제공하며, 각 모델의 특성이 다릅니다. 주요 모델의 특성과 적합한 용도를 이해하면 상황에 맞는 최적의 선택이 가능합니다.

Eleven Multilingual v2는 가장 고품질의 모델로, 29개 이상의 언어를 지원합니다. 감정 표현력이 뛰어나고 자연스러운 억양과 리듬을 생성합니다. 192kbps의 고품질 오디오를 출력하며, 유튜브 나레이션, 오디오북, 팟캐스트, 광고 등 최종 결과물의 품질이 중요한 작업에 적합합니다. 다만 지연 시간이 상대적으로 길고, Flash 모델에 비해 문자당 크레딧 소모가 높습니다.

Eleven Flash v2(또는 v2.5)는 속도에 최적화된 모델입니다. 지연 시간이 매우 짧아 실시간 대화나 빠른 테스트에 적합합니다. 같은 크레딧으로 Multilingual의 두 배 문자를 사용할 수 있습니다. 128kbps의 오디오 품질이며, 감정 표현력은 Multilingual보다 제한적입니다. 대화형 AI 에이전트, 빠른 프로토타이핑, 초안 확인 등에 적합합니다.

Eleven v3는 가장 최신의 모델로, 오디오 태그를 통한 감정 제어, 효과음 삽입, 다중 화자 대화 등의 고급 기능을 지원합니다. 이 모델에 대해서는 제18장에서 상세히 다루겠습니다.

Eleven English v1은 영어에 특화된 초기 모델입니다. 영어 발음과 억양의 정확도가 높지만, 다국어 지원은 제한적입니다.

모델 선택의 핵심 전략은 "용도에 맞는 모델을 선택한다"는 것입니다. 테스트나 초안 확인에는 Flash 모델을 사용하여 크레딧을 아끼고, 최종 결과물에만 Multilingual v2나 v3를 사용하는 것이 효율적입니다.

5.3 음성 설정 파라미터의 기본 이해

TTS 생성 시 조정할 수 있는 주요 파라미터가 있습니다. 이 파라미터들은 음성의 특성을 크게 좌우하므로, 기본적인 이해가 필요합니다. 각 파라미터에 대한 심층 분석은 제19장에서 다루겠습니다.

안정성(Stability)은 음성의 일관성과 변동성을 조절합니다. 안정성이 높으면 음성이 일관되고 예측 가능하지만 단조로울 수 있으며, 낮으면 더 표현력이 풍부하지만 때때로 예측 불가능한 결과가 나올 수 있습니다.

유사도 향상(Similarity Boost)은 생성된 음성이 원본 음성(복제된 음성의 경우)이나 선택한 음성과 얼마나 유사하게 발음하는지를 조절합니다.

스타일 강조(Style Exaggeration)는 음성의 감정적 표현과 스타일을 얼마나 강하게 적용할지를 조절합니다.

속도(Speed)는 생성된 음성의 재생 속도를 조절합니다. 기본값은 1.0이며, 0.7(느리게)에서 1.2(빠르게)까지 조절할 수 있습니다.

5.4 효과적인 TTS 텍스트 작성의 기본 원칙

TTS의 품질은 입력하는 텍스트의 품질에 크게 좌우됩니다. 같은 내용이라도 텍스트를 어떻게 작성하느냐에 따라 결과물의 자연스러움과 감정 표현이 달라집니다.

구두점을 적극 활용합니다. 쉼표는 자연스러운 호흡을, 마침표는 문장 종결의 확실한 멈춤을, 물음표는 울림 억양을, 느낌표는 강조와 에너지를 만들어냅니다. 말줄임표(...)는 생각하는 듯한 멈춤이나 망설임을 표현할 수 있습니다.

문장의 길이를 적절히 조절합니다. 너무 긴 문장은 부자연스러운 호흡과 억양을 만들 수 있습니다. 실제로 사람이 말하는 것처럼 적절한 길이로 문장을 나누는 것이 좋습니다.

텍스트를 구어체로 작성합니다. 딱딱한 문어체보다는 실제 대화에서 사용하는 자연스러운 구어체가 더 좋은 TTS 결과를 만들어냅니다. "본 보고서는 다음과 같은 내용을 포함합니다"보다 "이 보고서에서는 이런 내용들을 다루고 있습니다"가 더 자연스러운 음성을 생성합니다.

5.5 TTS 결과물의 후처리

생성된 음성 파일은 그 자체로도 충분히 활용 가능하지만, 후처리를 통해 품질을 한층 더 높일 수 있습니다. 무료 또는 저가의 오디오 편집 도구를 사용하여 불필요한 침묵 구간을 제거하거나, 배경 음악을 추가하거나, 여러 클립을 연결하는 등의 편집이 가능합니다.

특히 일레븐랩스의 효과음 생성 기능이나 Eleven Music 기능과 결합하면, TTS 음성 위에 배경 음악이나 효과음을 입히는 것도 모두 일레븐랩스 내에서 해결할 수 있습니다.

6 음성 복제(Voice Cloning) — 나만의 AI 음성 만들기

6.1 음성 복제의 개념

음성 복제는 실제 사람의 음성 샘플을 AI에게 학습시켜, 그 사람의 목소리 특성을 가진 AI 음성을 만드는 기능입니다. 복제된 음성으로 어떤 텍스트든 원하는 내용을 발화시킬 수 있습니다.

이 기능은 자신의 목소리를 복제하여 일관된 브랜드 음성을 만들거나, 전문 성우의 목소리를 복제하여 대량의 콘텐츠를 효율적으로 생성하거나, 다국어 콘텐츠에서 동일한 목소리를 유지하는 등의 용도로 활용됩니다.

6.2 인스턴트 클론(Instant Clone)과 프로페셔널 클론(Professional Voice Clone, PVC)

일레븐랩스는 두 가지 수준의 음성 복제를 제공합니다.

인스턴트 클론은 짧은 음성 샘플(약 1분 이상)만으로 빠르게 음성을 복제하는 기능입니다. 스타터 플랜 이상에서 사용 가능하며, 원본 음성의 대략적인 특성을 포착합니다. 빠르고 간편하지만, 원본과의 유사도에는 한계가 있을 수 있습니다. 무료 플랜에서는 인스턴트 클론을 사용할 수 없습니다.

프로페셔널 음성 복제(PVC)는 더 긴 음성 샘플(일반적으로 30분 이상 권장)을 사용하여 높은 정확도로 음성을 복제하는 기능입니다. 크리에이터 플랜 이상에서 사용 가능하며, 원본 음성의 미묘한 뉘앙스와 특성까지 정교하게 재현합니다.

6.3 좋은 음성 샘플 준비하기

음성 복제의 품질은 제공하는 음성 샘플의 품질에 크게 좌우됩니다. 좋은 샘플을 준비하기 위한 핵심 지침은 다음과 같습니다.

깨끗한 오디오를 확보합니다. 배경 소음, 에코, 다른 사람의 목소리가 없는 깨끗한 환경에서 녹음해야 합니다. 조용한 실내에서 외장 마이크를 사용하는 것이 이상적이지만, 스마트폰의 내장 마이크로도 조용한 환경이라면 충분한 품질을 얻을 수 있습니다.

다양한 감정과 톤을 포함합니다. 특히 인스턴트 클론의 경우, 샘플에 포함된 감정의 범위가 넓을수록 복제된 음성의 표현력이 풍부해집니다. 평범한 어조, 강조, 부드러운 말투, 약간의 감정 변화 등을 포함하는 것이 좋습니다.

자연스러운 발화를 녹음합니다. 문장을 또박또박 읽는 것보다, 자연스럽게 대화하듯이 말하는 것이 더 좋은 결과를 만듭니다. 실제로 누군가에게 이야기하는 것처럼 녹음하면 됩니다.

충분한 길이를 확보합니다. 인스턴트 클론은 최소 약 1분 이상이 권장되지만, 더 긴 샘플일수록 좋은 결과를 얻을 수 있습니다. PVC의 경우 30분 이상의 높은 품질 샘플이 권장됩니다.

6.4 음성 복제 시 주의사항

음성 복제 기능은 강력하지만, 몇 가지 중요한 주의사항이 있습니다.

반드시 본인의 음성이거나, 복제 권한을 가진 음성만 사용해야 합니다. 다른 사람의 목소리를 무단으로 복제하는 것은 일레븐랩스의 이용약관에 위배되며, 법적 문제를 야기할 수 있습니다.

일레븐랩스는 음성 복제 시 동의 확인 절차를 진행합니다. 복제하려는 음성이 본인의 것인지, 또는 해당 음성의 소유자로부터 명시적인 동의를 받았는지 확인하는 과정이 있습니다.

복제된 음성의 상업적 사용은 유료 플랜에서만 허용됩니다. 무료 플랜에서는 음성 복제 자체가 불가능하며, 스타터 플랜 이상에서 인스턴트 클론이, 크리에이터 플랜 이상에서 PVC가 가능합니다.

6.5 보이스 디자인(Voice Design)

음성 복제 외에도, 일레븐랩스에서는 보이스 디자인 기능을 통해 완전히 새로운 합성 음성을 만들 수 있습니다. 원하는 음성의 특성(성별, 연령대, 억양, 톤 등)을 텍스트로 설명하면, AI가 이에 맞는 새로운 음성을 생성합니다. 무료 플랜에서도 최대 3개의 커스텀 음성을 만들 수 있으므로, 음성 복제가 불가능한 무료 사용자에게는 좋은 대안이 될 수 있습니다.

7 보이스오버 스튜디오(Voiceover Studio) — 장편 콘텐츠 제작의 중심

7.1 보이스오버 스튜디오의 개요

보이스오버 스튜디오는 일레븐랩스에서 제공하는 타임라인 기반의 오디오 편집 환경입니다. 일반 TTS가 짧은 텍스트를 개별적으로 변환하는 데 적합하다면, 보이스오버 스튜디오는 긴 분량의 나레이션, 오디오북, 유튜브 영상 보이스오버 등 장편 콘텐츠 제작에 최적화되어 있습니다.

특히 주목할 만한 점은, 스튜디오 3.0 버전에서는 비디오 트랙과 자막 기능이 추가되어, 영상에 맞춰 보이스오버를 제작하는 것이 훨씬 수월해졌다는 것입니다. 음악과 효과음 레이어도 추가할 수 있어, 별도의 오디오 편집 소프트웨어 없이도 완성도 높은 오디오 콘텐츠를 만들 수 있습니다.

7.2 프로젝트 생성과 관리

보이스오버 스튜디오에서는 프로젝트 단위로 작업을 관리합니다. 무료 플랜에서는 최대 3개의 프로젝트를 만들 수 있습니다. 하나의 프로젝트 안에서 여러 섹션으로 나누어 작업할 수 있으며, 각 섹션마다 다른 음성을 지정하는 것도 가능합니다.

주목할 만한 점은 프로젝트를 생성하고 텍스트를 입력하는 것 자체는 크레딧을 소모하지 않는다는 것입니다. 크레딧은 실제로 음성을 생성(Convert)할 때만 소모됩니다. 따라서 프로젝트의 구조를 미리 충분히 기획하고, 텍스트를 다듬은 후에 생성 버튼을 누르면 불필요한 크레딧 소모를 방지할 수 있습니다.

7.3 다중 화자 프로젝트 관리

오디오 드라마, 대담 형식의 콘텐츠, 교육 영상 등에서는 여러 화자가 등장해야 하는 경우가 있습니다. 보이스오버 스튜디오에서는 텍스트의 각 부분에 서로 다른 음성을 지정할 수 있으므로, 하나의 프로젝트 안에서 여러 화자의 대화를 구현할 수 있습니다.

예를 들어 교육 콘텐츠에서 진행자의 설명 부분에는 A 음성을, 학생의 질문 부분에는 B 음성을 지정하면, 자연스러운 대화 형식의 오디오를 만들 수 있습니다.

7.4 발음 사전과의 연동

보이스오버 스튜디오에서는 발음 사전(Pronunciation Dictionary)을 업로드하여 사용할 수 있습니다. 프로젝트에 반복적으로 등장하는 고유 명사, 기술 용어, 브랜드명 등의 발음을 미리 지정해두면, 일관된 발음의 오디오를 생성할 수 있습니다. 이 기능의 상세한 활용법은 제21장에서 다루겠습니다.

7.5 효율적인 보이스오버 작업 워크플로우

보이스오버 스튜디오에서 크레딧을 절약하면서 효율적으로 작업하는 워크플로우는 다음과 같습니다.

첫 단계로 텍스트를 완전히 준비합니다. 프로젝트에 사용할 모든 텍스트를 미리 작성하고, 오타자 수정, 구두점 확인, 어색한 표현 수정 등을 완료합니다. 이 단계에서는 크레딧이 소모되지 않으므로 충분한 시간을 투자합니다.

두 번째 단계로 짧은 샘플을 먼저 테스트합니다. 전체 텍스트를 한꺼번에 변환하기 전에, 대표적인 짧은 구간만 먼저 변환하여 음성, 속도, 톤 등이 적절한지 확인합니다. 이 테스트에서 Flash 모델을 사용하면 크레딧을 절약할 수 있습니다.

세 번째 단계로 만족스러운 설정이 확인되면 전체 텍스트를 변환합니다. 이때는 최종 결과물에 사용할 모델(Multilingual v2 등)을 선택합니다.

네 번째 단계로 생성된 오디오를 검토하고 필요한 부분만 재생성합니다. 전체가 아닌 특정 섹션만 선택하여 다시 생성할 수 있으므로, 문제가 있는 부분만 수정하면 됩니다.

8.1 더빙 기능의 원리

일레븐랩스의 더빙 기능은 오디오나 비디오 콘텐츠를 다른 언어로 자동 번역하고, 원본 화자의 음성 특성을 유지하면서 번역된 텍스트를 음성으로 합성하는 기능입니다. 원본 콘텐츠의 화자를 자동으로 감지하고, 각 화자의 목소리를 복제하여 번역된 언어로 발화시킵니다.

이 기능은 29개 이상의 언어를 지원하며, 한국어로 제작된 콘텐츠를 영어나 일본어로 더빙하거나, 반대로 영어 콘텐츠를 한국어로 더빙하는 것이 가능합니다. 유튜브 크리에이터가 자신의 영상을 다국어로 제공하거나, 기업이 교육 자료를 글로벌 팀에게 배포할 때 매우 유용합니다.

8.2 자동 더빙과 더빙 스튜디오

일레븐랩스의 더빙은 두 가지 모드로 제공됩니다.

자동 더빙(Automatic Dubbing)은 파일을 업로드하고 대상 언어를 선택하면, AI가 자동으로 전체 과정을 처리하는 방식입니다. 빠르고 간편하지만, 세부적인 조정이 제한됩니다.

더빙 스튜디오(Dubbing Studio)는 더빙 과정을 세부적으로 제어할 수 있는 고급 인터페이스입니다. 번역된 텍스트를 직접 수정하고, 각 세그먼트의 타이밍을 조정하고, 특정 부분의 음성을 재생성하는 등 정밀한 편집이 가능합니다.

8.3 무료 플랜에서의 더빙 활용

더빙은 크레딧 소모가 상대적으로 높은 기능입니다. 무료 플랜에서는 크레딧의 제한으로 인해 짧은 분량의 콘텐츠만 더빙할 수 있습니다. 따라서 무료 플랜에서 더빙을 시도하려면, 가장 중요한 핵심 구간만 선별하여 더빙하는 전략이 필요합니다.

예를 들어, 10분짜리 유튜브 영상 전체를 더빙하기보다는, 영상의 핵심 1분짜리 하이라이트 클립만 더빙하여 해외 시청자를 끌어들이는 전략이 무료 플랜에서 현실적입니다.

8.4 더빙 품질을 높이는 팁

더빙 결과의 품질을 높이기 위한 핵심 팁은 다음과 같습니다. 원본 오디오의 품질이 높을수록 더빙 결과도 좋아집니다. 배경 소음이 적고 화자의 목소리가 선명한 원본 파일을 사용하는 것이 중요합니다. 여러 화자가 동시에 말하는 부분이 있으면 화자 구분이 어려워져 품질이 떨어질 수 있습니다. 더빙 스튜디오에서 번역된 텍스트를 검토하고 필요한 수정을 가하면, 최종 결과물의 정확도가 향상됩니다.

9 음성 변환(Voice Changer) — 목소리를 자유자재로

9.1 음성 변환 기능의 개요

음성 변환(Voice Changer, 이전 명칭 Speech-to-Speech)은 녹음된 음성을 다른 목소리로 변환하는 기능입니다. 원본 음성의 어조, 감정, 리듬은 유지하면서, 목소리의 특성만 다른 음성으로 바꿉니다.

예를 들어, 자신의 목소리로 녹음한 나레이션을 보이스 라이브러리에 있는 전문 나레이터 풍의 음성으로 변환하거나, 여러 명의 대화를 혼자서 녹음한 뒤 각 화자마다 다른 음성으로 변환하는 것이 가능합니다.

9.2 음성 변환의 장점과 활용

음성 변환이 순수 TTS에 비해 가지는 고유한 장점이 있습니다. 텍스트만으로는 전달하기 어려운 미묘한 감정과 뉘앙스를 직접 녹음을 통해 입력할 수 있다는 점입니다. 풍자적인 어조, 특정 단어에 대한 강조, 말의 빠르기 변화 등 실제 연기에 가까운 요소를 본인의 녹음에 담고, 이를 원하는 목소리로 변환하면 TTS보다 훨씬 섬세한 결과를 얻을 수 있습니다.

또한 프라이버시 측면에서도 유용합니다. 자신의 목소리를 직접 공개하고 싶지 않은 크리에이터가, 자신의 자연스러운 연기를 담은 녹음을 다른 음성으로 변환하여 사용할 수 있습니다.

9.3 무료 플랜에서의 음성 변환

음성 변환은 분(minutes) 단위로 크레딧이 소모됩니다. 무료 플랜에서의 사용 가능 분량은 제한적이므로, 짧은 구간에 대해서만 전략적으로 사용하는 것이 좋습니다. 전체 10분짜리 나레이션을 변환하기보다는, 가장 핵심적인 30초짜리 클립에 대해서만 변환을 수행하는 방식이 현실적입니다.

10 음성 분리(Voice Isolator) — 깨끗한 음성 추출

10.1 음성 분리의 원리와 용도

음성 분리(Voice Isolator)는 AI를 사용하여 오디오 파일에서 음성만 깨끗하게 추출하고, 배경 소음(에코, 마이크 피드백, 거리 소음, 음악 등)을 제거하는 기능입니다.

이 기능은 팟캐스트 녹음에서 배경 잡음을 제거하거나, 인터뷰 영상에서 음성만 깔끔하게 추출하거나, 음성 복제용 샘플에서 불필요한 소음을 제거하여 복제 품질을 높이거나, 기존 녹음의 음질을 향상시키는 등의 용도로 활용됩니다.

10.2 활용 전략

음성 분리는 다른 기능과 결합할 때 특히 강력합니다. 예를 들어, 소음이 있는 환경에서 녹음한 음성을 먼저 Voice Isolator로 정리한 후, 정리된 음성을 인스턴트 클론의 샘플로 사용하면 복제 품질이 크게 향상됩니다. 또한 팟캐스트 녹음 후 후처리 단계에서 배경 소음을 제거하면, 전문 스튜디오에서 녹음한 것과 같은 깨끗한 오디오를 얻을 수 있습니다.

11 효과음 생성(Sound Effects) — AI로 만드는 사운드

11.1 효과음 생성의 개요

효과음 생성은 텍스트 프롬프트를 입력하면 AI가 해당하는 효과음을 생성하는 기능입니다. "빗소리", "도시의 교통 소음", "칼을 부딪치는 소리", "관중의 박수 소리" 등 자연어로 원하는 효과음을 설명하면, AI가 이에 맞는 오디오를 만들어냅니다.

이 기능은 생성 횟수(generations) 단위로 크레딧이 소모됩니다. 영상 편집, 게임 개발, 팟캐스트 제작, 프레젠테이션 등에서 별도의 효과음 라이브러리를 구매하지 않고도 필요한 사운드를 만들 수 있어 매우 실용적입니다.

11.2 효과적인 효과음 프롬프트 작성

좋은 효과음을 얻기 위한 프롬프트 작성 기법을 정리하면 다음과 같습니다. 구체적으로 묘사합니다. "비 소리"보다는 "창문에 부딪히는 세찬 빗소리, 먼 곳에서 들리는 천둥"이 더 정확한 결과를 만듭니다. 원하는 분위기를 함께 설명합니다. "평화로운", "긴장감 있는", "활기찬" 등의 분위기 키워드를 추가하면 결과물의 톤이 달라집니다. 복잡한 효과음은 여러 개의 단순한 효과음으로 나누어 생성하고, 오디오 편집 소프트웨어에서 레이어링하는 것이 더 좋은 결과를 얻는 방법입니다.

12 AI 음악 생성(Eleven Music) — 텍스트로 작곡하기

12.1 Eleven Music의 개요

Eleven Music은 2025년 8월에 출시된 기능으로, 자연어 프롬프트를 입력하면 스튜디오급 품질의 음악을 생성하는 AI 음악 생성 도구입니다. 장르, 분위기, 템포, 악기 구성 등을 텍스트로 설명하면, 이에 맞는 완성된 음악 트랙이 생성됩니다.

특히 중요한 점은, 유료 플랜에서 생성된 음악은 상업적 사용이 허가된 로열티 프리 음악이라는 것입니다. 일레븐랩스는 Merlin 등 음악 산업의 주요 파트너와 라이선스 계약을 체결하여, 생성된 음악의 상업적 사용 권한을 보장합니다.

12.2 음악 프롬프트 작성 팁

효과적인 음악 프롬프트를 작성하기 위해서는 장르를 명시하고(팝, 재즈, 일렉트로닉, 클래식 등), 분위기를 설명하며(에너지 넘치는, 잔잔한, 몽환적인, 긴장감 있는 등), 원하는 악기를 지정하고(피아노, 어쿠스틱 기타, 일렉트릭 베이스, 드럼 등), 템포를 제시하는(빠른, 느린, 중간 등) 것이 효과적입니다.

프롬프트가 구체적일수록 원하는 결과에 가까운 음악이 생성됩니다. 또한 섹션 편집 기능을 통해 특정 구간을 수정하거나 보컬을 추가하는 등의 세부 조정도 가능합니다.

12.3 활용 시나리오

Eleven Music은 유튜브 영상의 배경 음악, 팟캐스트의 인트로와 아웃트로 음악, 광고나 프레젠테이션의 배경 음악, 소셜 미디어 콘텐츠의 사운드트랙 등 다양한 용도로 활용할 수 있습니다. TTS로 생성한 나레이션과 Eleven Music으로 생성한 배경 음악, 효과음 생성으로 만든 사운드 효과를 결합하면, 일레븐랩스 하나로 완성도 높은 오디오 콘텐츠를 만들 수 있습니다.

13 음성 인식(Speech-to-Text) — AI 전사의 새로운 기준

13.1 Scribe 모델의 특징

일레븐랩스의 음성 인식 기능은 자체 개발한 Scribe 모델을 기반으로 합니다. 90개 이상의 언어를 지원하며, 화자 분리(Speaker Diarization), 단어 수준의 타임스탬프, 높은 정확도의 전사를 제공합니다.

이 기능은 API와 웹 UI 두 가지 방식으로 제공됩니다. API를 통한 사용은 대규모 자동화된 전사 작업에 적합하고, 웹 UI를 통한 사용은 수동으로 파일을 업로드하여 전사 결과를 확인하고 편집하는 데 적합합니다.

13.2 무료 플랜에서의 음성 인식 활용

무료 플랜에서도 음성 인식 기능에 접근할 수 있지만, 사용 가능한 시간이 제한됩니다. API를 통한 사용은 약 2.5시간, UI를 통한 사용은 약 12분 수준의 전사가 가능한 것으로 알려져 있습니다.

제한된 쿼터를 효율적으로 사용하려면, 전사가 꼭 필요한 핵심 오디오만 선별하여 사용하는 것이 중요합니다. 전체 회의 녹음을 전사하기보다는, 중요한 의사 결정 구간이나 핵심 발표 부분만 전사하는 전략이 효율적입니다.

13.3 전사 결과의 활용

전사된 텍스트는 다양한 방식으로 활용할 수 있습니다. 회의 내용의 문서화, 인터뷰 내용의 기록, 영상 콘텐츠의 자막 생성, 팟캐스트의 텍스트 버전 생성 등이 대표적입니다. 또한 전사된 텍스트를 일레븐랩스의 TTS로 다시 변환하면, 원본 오디오를 다른 목소리나 다른 언어로 재생성하는 것도 가능합니다.

14 대화형 AI 에이전트(Conversational AI) — 음성 AI 비서 구축

14.1 대화형 AI의 개념

일레븐랩스의 대화형 AI(Conversational AI)는 실시간 음성 기반의 AI 에이전트를 구축할 수 있는 플랫폼입니다. 사용자가 음성으로 질문하면 AI가 음성으로 답변하는, 자연스러운 대화형 인터페이스를 만들 수 있습니다.

이 기능은 고객 서비스 자동화, 교육용 인터랙티브 튜터, 가상 안내원, 인터랙티브 캐릭터 등에 활용될 수 있습니다. 31개 이상의 언어를 지원하며, 대화 중 적절한 타이밍에 말을 끊거나 기다리는 등의 턴테이킹(Turn-Taking) 기술도 적용되어 있습니다.

14.2 에이전트 구축의 기본 요소

대화형 AI 에이전트를 구축하기 위해서는 몇 가지 핵심 요소를 설정해야 합니다. 에이전트의 역할과 성격을 정의하는 시스템 프롬프트, 에이전트가 사용할 음성 선택, 지원하는 언어 설정, 대화의 컨텍스트와 지식 기반 등을 구성합니다.

일레븐랩스의 공식 문서에서는 에이전트 프롬프트 작성에 대한 구체적인 가이드를 제공하고 있습니다. 지시를 깔끔한 섹션으로 구분하고, 가능한 한 간결하게 작성하며, 핵심적인 지시를 강조하는 것이 좋은 프롬프트의 기본 원칙입니다.

14.3 크레딧 소모와 관리

대화형 AI는 분(minutes) 단위로 크레딧이 소모됩니다. 또한 텍스트 메시지 전송 건수도 별도로 측정됩니다. 무료 플랜에서는 사용 가능한 대화 시간이 매우 제한적이므로, 프로토타이핑이나 테스트 목적으로만 활용하는 것이 현실적입니다. 본격적인 대화형 AI 서비스를 운영하려면 유료 플랜이 필요합니다.

15.1 오디오 네이티브의 개요

오디오 네이티브(Audio Native)는 웹페이지에 임베드할 수 있는 오디오 플레이어를 제공하는 기능입니다. 블로그 글, 뉴스 기사, 뉴스레터 등의 텍스트 콘텐츠를 자동으로 음성으로 변환하여, 독자가 "읽기" 대신 "듣기"로 콘텐츠를 소비할 수 있게 해줍니다.

이 기능은 웹사이트에 간단한 코드를 삽입하는 것만으로 설정할 수 있으며, 페이지의 텍스트 내용을 자동으로 감지하여 음성으로 변환합니다. 콘텐츠가 업데이트되면 오디오도 자동으로 갱신됩니다.

15.2 활용 사례

오디오 네이티브는 블로그 운영자가 접근성을 높이고 싶을 때, 뉴스 매체가 오디오 뉴스를 제공하고 싶을 때, 교육 플랫폼이 학습 자료의 오디오 버전을 제공하고 싶을 때, 뉴스레터 발행자가 구독자에게 듣기 옵션을 제공하고 싶을 때 유용합니다. 독자의 콘텐츠 소비 경험을 다양화하고, 이동 중이거나 눈을 사용하기 어려운 상황에서도 콘텐츠에 접근할 수 있게 해줍니다.

16.1 보이스 라이브러리의 구조

보이스 라이브러리는 일레븐랩스 커뮤니티가 공유한 음성을 탐색하고 사용할 수 있는 마켓플레이스입니다. 수천 개의 다양한 음성이 등록되어 있으며, 성별, 연령대, 언어, 억양, 용도 등 다양한 필터로 검색할 수 있습니다.

보이스 라이브러리에 있는 모든 음성은 무료 상업적 사용 라이선스와 함께 제공됩니다. 이는 무료 플랜 사용자에게도 적용되는 것이 아니라, 유료 플랜에서 상업적 사용 권한이 부여된 사용자가 해당 음성을 상업적으로 활용할 수 있다는 의미입니다.

검색 기능은 키워드 검색뿐만 아니라, 오디오 파일을 업로드하여 유사한 음성을 찾는 것도 가능합니다. 원하는 음성 스타일의 샘플이 있다면, 이를 업로드하여 비슷한 음성을 보이스 라이브러리에서 찾을 수 있습니다.

16.2 아이코닉 마켓플레이스(Iconic Marketplace)

아이코닉 마켓플레이스는 유명인의 공인된 AI 음성을 라이선스하여 사용할 수 있는 프리미엄 마켓플레이스입니다. 이 음성들은 해당 인물의 직접적인 동의를 받아 제작된 것으로, 합법적이고 윤리적으로 사용할 수 있습니다.

이 기능은 콘텐츠에 특별한 캐릭터와 매력을 부여하고 싶을 때, 역사적 인물이나 문화적 아이콘의 목소리를 교육이나 엔터테인먼트에 활용하고 싶을 때 유용합니다. 다만 아이코닉 마켓플레이스의 음성은 별도의 라이선스 비용이 발생할 수 있으므로, 사용 전 가격과 이용 조건을 확인해야 합니다.

16.3 무료 플랜에서의 보이스 라이브러리 활용 전략

무료 플랜에서도 보이스 라이브러리의 음성을 탐색하고, "My Voices"에 저장하여 TTS에 사용할 수 있습니다. 무료 플랜에서는 음성 복제가 불가능하므로, 보이스 라이브러리의 다양한 음성이 사실상 유일한 음성 다양화 수단입니다.

자신의 콘텐츠에 적합한 음성을 미리 충분히 탐색하고, 최적의 음성을 선정하여 저장해두면, 실제 작업 시 음성 선택에 소비하는 시간을 절약할 수 있습니다. 또한 동일한 프로젝트 내에서 일관된 음성을 사용하는 것이 전문적인 인상을 주므로, 한번 선정한 음성은 프로젝트 전체에서 일관되게 사용하는 것이 좋습니다.

제3부. 품질 극대화를 위한 고급 전략

17 TTS 품질을 극대화하는 텍스트 작성 기법

17.1 텍스트가 음성 품질을 결정한다

TTS의 결과물 품질에 가장 큰 영향을 미치는 요소는 입력 텍스트의 품질입니다. 같은 음성, 같은 모델, 같은 설정을 사용하더라도 텍스트를 어떻게 작성하느냐에 따라 결과물의 자연스러움, 감정 표현, 명확성이 크게 달라집니다.

이것은 크레딧 절약과도 직결됩니다. 잘 작성된 텍스트는 첫 번째 시도에서 만족스러운 결과를 만들어내므로, 재생성의 필요를 줄여 크레딧을 절약할 수 있습니다.

17.2 구두점을 통한 리듬 제어

구두점은 TTS 엔진에게 보내는 가장 기본적인 "연출 지시"입니다.

쉼표(,)는 짧은 호흡과 자연스러운 멈춤을 만듭니다. 긴 문장 안에서 적절한 위치에 쉼표를 넣으면, AI가 문장을 자연스러운 단위로 나누어 발화합니다.

마침표(.)는 확실한 문장 종결을 표시합니다. 선언적이고 확신에 찬 어조를 만들어냅니다.

물음표(?)는 올림 억양을 생성합니다. 질문의 의도를 명확하게 전달할 수 있습니다.

느낌표(!)는 강조와 에너지를 부여합니다. 다만 과도하게 사용하면 부자연스러워질 수 있으므로, 정말 강조가 필요한 곳에서만 사용하는 것이 좋습니다.

말줄임표(...)는 생각하는 듯한 멈춤, 망설임, 또는 의미심장한 여운을 만들어냅니다. Eleven v3 모델에서는 특히 효과적으로, 멈춤의 길이와 분위기에 영향을 줍니다.

대시(—)는 갑작스러운 전환이나 짧은 끊김을 표현할 수 있습니다.

17.3 서사적 스타일의 텍스트 작성

일레븐랩스의 공식 가이드에서는 텍스트를 "서사적 스타일(Narrative Styling)"로 작성할 것을 권장합니다. 이는 텍스트를 마치 대본이나 소설처럼, 톤과 감정이 자연스럽게 흐르도록 작성하는 것을 의미합니다.

건조한 정보 전달 스타일의 "본 제품은 2026년 3월 출시 예정입니다. 가격은 미정입니다." 보다는, 자연스러운 대화 스타일의 "이 제품은, 2026년 3월에 출시될 예정인데요. 가격은 아직 확정되지 않았습니니다."가 더 자연스러운 음성을 생성합니다.

핵심은 "사람이 실제로 이렇게 말할까?"를 항상 자문하면서 텍스트를 작성하는 것입니다. 문어체와 구어체의 차이를 인식하고, TTS용 텍스트는 구어체에 가깝게 작성하는 것이 좋습니다.

17.4 대문자를 통한 강조 (영어 텍스트의 경우)

영어 텍스트에서는 특정 단어를 대문자로 표기하면 강조 효과를 줄 수 있습니다. "This is REALLY important."에서 REALLY는 더 강하게 발화됩니다. 이는 특히 Eleven v3 모델에서 효과적입니다.

한국어의 경우에는 대문자 개념이 없으므로, 강조를 위해서는 다른 기법을 사용해야 합니다. 강조하고 싶은 단어 앞뒤에 쉼표를 넣어 자연스러운 멈춤을 만들거나, 느낌표를 활용하거나, Eleven v3의 오디오 태그를 사용하는 방법이 있습니다.

17.5 문장 길이의 최적화

TTS에서 문장의 길이는 결과물의 자연스러움에 큰 영향을 미칩니다. 너무 긴 문장은 AI가 부자연스러운 호흡 패턴이나 단조로운 억양을 생성하게 만들 수 있습니다.

이상적인 문장 길이는 실제로 한 번의 호흡에 자연스럽게 발화할 수 있는 정도입니다. 일반적으로 30자에서 60자 내외가 TTS에 적합한 문장 길이입니다. 원본 텍스트가 긴 문장으로 구성되어 있다면, TTS 변환 전에 적절한 길이로 분할하는 것이 좋습니다.

18 Eleven v3 모델의 오디오 태그와 감정 표현 마스터하기

18.1 Eleven v3의 혁신적인 특성

Eleven v3는 일레븐랩스의 최신 TTS 모델로, 오디오 태그를 통한 감정과 비언어적 표현 제어라는 혁신적인 기능을 도입했습니다. 기존 모델에서는 텍스트의 내용과 구두점으로만 간접적으로 감정을 유도할 수 있었다면, v3에서는 대괄호 안에 명시적인 지시를 넣어 원하는 감정이나 효과를 직접 제어할 수 있습니다.

18.2 음성 관련 오디오 태그

v3에서 사용할 수 있는 주요 음성 관련 오디오 태그는 다음과 같습니다.

웃음 표현으로 [laughs], [laughs harder], [starts laughing], [wheezing] 등이 있습니다. 속삭임은 [whispers]로 표현합니다. 한숨이나 호흡은 [sighs], [exhales]로 표현합니다. 감정 표현으로 [sarcastic], [curious], [excited], [crying], [snorts], [mischievously] 등이 사용 가능합니다.

이 태그들을 텍스트 안에 자연스럽게 삽입하면, AI가 해당 위치에서 지정된 감정이나 효과를 음성에 반영합니다. 예를 들어 "정말요? [laughs] 그건 정말 재미있네요!"라고 입력하면, "정말요?" 다음에 웃음 소리가 삽입된 후 나머지 문장이 발화됩니다.

18.3 효과음 태그

환경음이나 효과음을 직접 삽입할 수 있는 태그도 있습니다. [gunshot], [applause], [clapping], [explosion] 등이 사용 가능합니다. 이 태그들은 오디오 드라마, 인터랙티브 콘텐츠, 게임 등에서 특히 유용합니다.

18.4 실험적 태그

좀 더 실험적인 태그로는 [strong X accent](X를 원하는 억양으로 대체), [sings], [woo], [fart] 등이 있습니다. 이러한 실험적 태그는 음성에 따라 결과가 일관적이지 않을 수 있으므로, 실제 사용 전에 충분히 테스트하는 것이 좋습니다.

18.5 태그 활용의 핵심 원칙

오디오 태그를 효과적으로 사용하기 위한 핵심 원칙이 있습니다.

음성 선택이 가장 중요합니다. 태그가 제대로 작동하려면, 선택한 음성이 해당 감정이나 효과를 표현할 수 있는 범위를 가지고 있어야 합니다. 차분한 내레이션 전용 음성에 [shout] 태그를 사용해도 효과가 미미할 수 있습니다.

안정성(Stability) 설정과 함께 사용합니다. v3에서 안정성 설정은 "Creative", "Natural", "Robust" 세 가지로 나뉩니다. 오디오 태그의 표현력을 극대화하려면 Creative 또는 Natural 설정을 사용하는 것이 좋습니다. Robust 설정은 안정적이지만 태그에 대한 반응이 약합니다.

태그를 조합하여 사용할 수 있습니다. 여러 태그를 결합하면 더 복잡한 감정 표현이 가능합니다. 다만, 지나치게 많은 태그를 사용하면 결과가 예측 불가능해질 수 있으므로 적절한 수준을 유지해야 합니다.

18.6 Enhance 기능의 활용

일레븐랩스 UI에서는 "Enhance" 버튼을 클릭하면, 입력한 텍스트에 자동으로 적절한 오디오 태그를 추가해주는 기능이 있습니다. 이 기능은 내부적으로 LLM(대규모 언어 모델)을 사용하여 텍스트의 맥락을 분석하고, 적합한 감정 태그와 비언어적 표현을 자동으로 삽입합니다. 오디오 태그를 직접 작성하는 데 익숙하지 않은 사용자에게 좋은 출발점이 됩니다.

19 음성 설정 파라미터 최적화 — 안정성, 유사도, 스타일, 속도

19.1 안정성(Stability) 슬라이더의 심층 이해

안정성 슬라이더는 음성 출력의 일관성과 표현 범위를 조절하는 가장 중요한 파라미터입니다.

안정성을 높게 설정하면 음성이 매우 일관되고 예측 가능합니다. 매번 생성해도 거의 같은 톤과 리듬을 유지합니다. 뉴스 앵커 스타일의 균일한 내레이션, 기업 교육 자료 등에 적합

합니다. 다만 감정 표현이 제한되고 단조롭게 들릴 수 있습니다.

안정성을 낮게 설정하면 음성의 변동성이 커지고 더 풍부한 감정 표현이 가능합니다. 오디오 오북, 이야기 나레이션, 감성적인 광고 등 감정 전달이 중요한 콘텐츠에 적합합니다. 다만 때때로 예측 불가능한 결과가 나올 수 있으며, 오디오 아티팩트(부자연스러운 소리)가 발생할 가능성이 있습니다.

Eleven v3 모델에서는 안정성이 "Creative"(가장 표현적), "Natural"(균형), "Robust"(가장 안정적) 세 단계로 구분됩니다. 대부분의 일반적인 용도에서는 Natural 설정이 적합하며, 감정이 풍부해야 하는 콘텐츠에서는 Creative를, 일관성이 가장 중요한 경우에는 Robust를 선택합니다.

19.2 유사도 향상(Similarity Boost) 슬라이더

유사도 향상은 생성된 음성이 선택한 음성(또는 복제된 음성)과 얼마나 유사하게 들리는지를 조절합니다. 유사도를 높이면 원본 음성에 더 가까워지지만, 너무 높게 설정하면 오디오 아티팩트가 발생할 수 있습니다. 너무 낮게 설정하면 원본 음성과의 유사성이 떨어집니다.

음성 복제를 사용하는 경우, 복제된 음성이 원본과 최대한 유사하게 들리도록 유사도를 높이고 싶을 수 있지만, 지나치게 높은 설정은 오히려 품질을 저하시킬 수 있습니다. 중간 정도의 설정에서 시작하여 결과를 들어보며 조금씩 조정하는 것이 좋습니다.

19.3 속도(Speed) 조절

속도 설정은 생성된 음성의 재생 속도를 조절합니다. 기본값은 1.0이며, 최소 0.7(30% 느리게)에서 최대 1.2(20% 빠르게)까지 조절할 수 있습니다.

지나치게 극단적인 값은 음성 품질에 영향을 줄 수 있으므로 주의가 필요합니다. 또한 속도 조절은 음성 생성 후에 적용하는 것보다, 처음 생성할 때 적절한 속도를 설정하는 것이 품질 면에서 더 나은 결과를 만듭니다.

일레븐랩스의 공식 가이드에서는 느린 속도를 기본으로 생성하고, 필요하면 후처리에서 빠르게 만드는 것이 그 반대보다 낫다고 조언합니다. 느린 음성을 5% 이상 빠르게 만드는 것보다, 빠른 음성을 5% 이상 느리게 만드는 것이 품질 저하가 더 크기 때문입니다.

19.4 파라미터 설정 조합의 실전 가이드

용도에 따른 파라미터 설정 조합을 참고하면 다음과 같습니다.

유튜브 나레이션에는 안정성을 중간 수준(Natural)으로, 유사도를 중간에서 약간 높게, 속도를 0.95에서 1.0으로 설정하는 것이 일반적으로 좋은 결과를 만듭니다.

오디오북에는 안정성을 중간에서 약간 낮게(Creative와 Natural 사이), 유사도를 중간 수준으로, 속도를 0.9에서 0.95로 설정하는 것이 감정 표현과 청취 편의성의 균형을 맞춥니다.

교육 자료에는 안정성을 높게(Robust 또는 Natural), 유사도를 중간 수준으로, 속도를 0.9에서 0.95로 설정하는 것이 명확하고 일관된 발음을 보장합니다.

광고 보이스오버에는 안정성을 낮게에서 중간(Creative), 유사도를 높게, 속도를 1.0에서 1.1로 설정하는 것이 에너지 넘치고 매력적인 음성을 만들어냅니다.

20 텍스트 정규화(Text Normalization) — 숫자, 약어, 기호 처리법

20.1 TTS에서 텍스트 정규화가 중요한 이유

TTS 모델은 일반적인 텍스트를 자연스럽게 읽지만, 숫자, 약어, 기호, 전화번호, 날짜, URL 등의 특수한 형식에서는 부자연스러운 발음이 나올 수 있습니다. 예를 들어 "\$1,000,000"을 "일 달러 콤마 영영영 콤마 영영영"으로 읽거나, "2026-03-31"을 "이공이육 대시 공삼 대시 삼일"로 읽을 수 있습니다.

텍스트 정규화란 이러한 특수 형식을 TTS가 자연스럽게 읽을 수 있는 형태로 변환하는 것을 말합니다.

20.2 일레븐랩스의 자동 정규화 기능

일레븐랩스는 기본적으로 모든 TTS 모델에 대해 자동 정규화 기능을 활성화하고 있습니다. 이 기능은 숫자, 날짜 등의 일반적인 형식을 자동으로 자연스러운 발음으로 변환합니다.

다만 모델에 따라 정규화의 정확도가 다릅니다. Multilingual v2 모델은 대형 모델이므로 숫자와 구문의 자연스러운 발음에 더 강하지만, Flash 모델은 경량 모델이므로 일부 복잡한 형식에서 부자연스러운 발음이 나올 수 있습니다.

20.3 수동 정규화 기법

자동 정규화로 처리되지 않는 경우, 수동으로 텍스트를 정규화할 수 있습니다. 핵심 원칙은 "사람이 소리 내어 읽을 때의 형태로 텍스트를 변환한다"는 것입니다.

숫자의 경우, "123"을 "백이십삼"으로, 전화번호 "010-1234-5678"을 "공일공, 일이삼사, 오육칠팔"로 변환합니다. 통화 기호의 경우 "\$45.67"을 "45달러 67센트"로 변환합니다. 날

짜의 경우 "2026-03-31"을 "2026년 3월 31일"로 변환합니다. 약어의 경우 "AI"를 "에이아이"로, "URL"을 "유알엘"로 변환합니다. 단위의 경우 "100km"를 "백 킬로미터"로, "25%"를 "이십오 퍼센트"로 변환합니다.

20.4 LLM을 활용한 자동 정규화

대량의 텍스트를 정규화해야 하는 경우, LLM(대규모 언어 모델, 예: ChatGPT)에게 정규화 작업을 맡기는 것도 효율적인 방법입니다. "다음 텍스트를 TTS 엔진이 자연스럽게 읽을 수 있도록 정규화해줘. 숫자는 한글로 풀어쓰고, 약어는 전체 형태로 확장하고, 기호는 한국어로 표기해줘."라고 지시하면, 대부분의 정규화가 자동으로 처리됩니다.

21 발음 사전(Pronunciation Dictionary) 활용법

21.1 발음 사전의 개념

발음 사전은 특정 단어나 구문의 발음을 미리 지정해두는 기능입니다. AI가 특정 단어를 잘못 발음하거나, 고유 명사나 기술 용어의 발음이 일관되지 않을 때, 발음 사전에 올바른 발음을 등록해두면 항상 지정된 발음으로 읽게 됩니다.

이 기능은 보이스오버 스튜디오와 더빙 스튜디오, 그리고 API를 통한 음성 합성에서 사용할 수 있습니다. 발음 사전은 TXT 파일 또는 PLS(Pronunciation Lexicon Specification) 형식으로 업로드합니다.

21.2 발음 지정 방식

발음을 지정하는 방식은 크게 두 가지입니다.

음소 태그(Phoneme Tag)는 CMU Arpabet이나 IPA(국제 음성 기호)를 사용하여 정확한 발음을 음소 단위로 지정하는 방법입니다. 가장 정밀한 발음 제어가 가능하지만, 음소 기호에 대한 지식이 필요합니다. 현재 음소 태그는 Eleven Flash v2와 Eleven English v1 모델에서만 호환됩니다.

별칭 태그(Alias Tag)는 특정 단어를 다른 단어나 구문으로 대체하여 발음하도록 지정하는 방법입니다. 예를 들어 "AI"를 만날 때마다 "인공지능"으로 읽도록 지정하거나, 복잡한 외래어를 더 발음하기 쉬운 표기로 대체할 수 있습니다. 음소 태그를 지원하지 않는 Multilingual v2 모델에서도 사용 가능합니다.

21.3 실전 활용 사례

발음 사전이 특히 유용한 경우를 정리하면 다음과 같습니다. 브랜드명이나 제품명의 일관된 발음 보장, 인명(특히 외국인 이름)의 정확한 발음, 기술 용어나 학술 용어의 올바른 발음, 약어의 읽는 방식 통일(글자별로 읽기 vs 단어로 읽기) 등에 활용됩니다.

한 번 발음 사전을 만들어두면 프로젝트 전체에 걸쳐 일관된 발음이 유지되므로, 장편 콘텐츠나 시리즈물 제작에서 특히 가치가 높습니다.

제4부. 크레딧 관리와 비용 최적화 전략

22 크레딧 소모 패턴 분석과 최적 배분 전략

22.1 기능별 크레딧 소모 체계 이해

일레븐랩스의 각 기능은 서로 다른 단위와 비율로 크레딧을 소모합니다. 이를 정확히 이해하는 것이 크레딧 최적화의 기초입니다.

TTS는 문자 수(characters) 단위로 소모됩니다. 입력한 텍스트의 문자 수가 곧 크레딧 소모량입니다. 1,000자를 입력하면 1,000자의 크레딧이 소모됩니다.

더빙과 음성 변환, 대화형 AI는 분(minutes) 단위로 소모됩니다. 음성 인식(STT)은 시간(hours) 단위로 소모됩니다. 효과음은 생성 횟수(generations) 단위로 소모됩니다.

중요한 점은, 이 모든 기능의 크레딧이 별도의 풀에서 관리된다는 것입니다. 즉, TTS의 문자 크레딧과 더빙의 분 크레딧은 서로 다른 풀입니다. 각 플랜에서 기능별로 제공되는 크레딧 양이 다르므로, 자신이 주로 사용하는 기능에 맞는 플랜을 선택하는 것이 중요합니다.

22.2 자신의 사용 패턴 파악하기

크레딧 최적화의 첫 걸음은 자신이 어떤 기능을 얼마나 사용하는지 파악하는 것입니다. 첫 한 달 동안 각 기능의 사용량을 간단히 기록해봅니다.

TTS에서 한 달에 대략 몇 문자를 생성하는지, 더빙이나 음성 변환을 얼마나 사용하는지, 효과음이나 음악 생성을 얼마나 활용하는지, 크레딧이 월 중간에 소진되는지 아니면 월말에도 여유가 있는지 등을 관찰합니다.

이 데이터를 바탕으로, 크레딧이 불필요하게 소비되는 지점을 식별하고, 더 효율적인 사용 방법을 찾을 수 있습니다.

22.3 크레딧 절약의 핵심 원칙

기능에 관계없이 적용할 수 있는 크레딧 절약의 핵심 원칙을 정리하면 다음과 같습니다.

텍스트를 완벽히 준비한 후 생성합니다. TTS의 가장 큰 크레딧 낭비는 오타자나 어색한 표현 때문에 재생성하는 것입니다. 텍스트를 충분히 검토하고 다듬은 후에 생성 버튼을 누릅니다.

테스트에는 Flash 모델을 사용합니다. 초안 확인이나 방향성 테스트에는 크레딧 소모가 적은 Flash 모델을 사용하고, 최종 결과물에만 Multilingual 모델을 사용합니다.

짧은 구간부터 테스트합니다. 긴 텍스트를 한꺼번에 변환하기 전에, 대표적인 짧은 구간만 먼저 변환하여 설정이 적절한지 확인합니다.

이전 생성물을 재활용합니다. 이미 생성한 음성이 필요하다면, 다시 생성하지 않고 이력에서 다운로드합니다.

불필요한 재생성을 줄입니다. 아주 사소한 차이에 대해 반복적으로 재생성하면 크레딧이 빠르게 소진됩니다. "충분히 좋은" 수준에서 만족하는 판단력도 필요합니다.

23 Flash 모델 vs Multilingual v2 — 상황별 현명한 선택

23.1 두 모델의 핵심 차이 정리

Flash 모델과 Multilingual v2 모델의 차이를 다시 한번 명확히 정리합니다.

Flash 모델은 지연 시간이 가장 짧고, 문자 소모가 적어(동일 플랜에서 Multilingual의 2배 분량 사용 가능) 효율적이며, 128kbps 오디오 품질을 제공합니다. 제한된 언어 지원과 상대적으로 낮은 감정 표현력이 단점입니다.

Multilingual v2는 29개 이상의 언어를 지원하고, 뛰어난 감정 표현력과 자연스러운 억양을 제공하며, 192kbps 고품질 오디오를 출력합니다. 다만 지연 시간이 더 길고, 같은 내용에 더 많은 문자 크레딧을 소모합니다.

23.2 상황별 최적 모델 선택

대화형 AI 에이전트에는 Flash 모델이 적합합니다. 실시간 대화에서는 응답 속도가 핵심이기 때문입니다.

유튜브 나레이션, 오디오북, 팟캐스트에는 Multilingual v2가 적합합니다. 최종 결과물의 품질이 중요하고, 실시간 응답이 필요하지 않기 때문입니다.

내부 테스트나 프로토타이핑에는 Flash 모델이 적합합니다. 크레딧을 절약하면서 방향성을 확인하는 데 충분합니다.

다국어 콘텐츠에는 Multilingual v2가 적합합니다. Flash 모델의 언어 지원은 제한적이므로, 다국어 콘텐츠에서는 Multilingual이 필수적입니다.

23.3 하이브리드 워크플로우

가장 효율적인 접근은 하이브리드 워크플로우입니다. 텍스트 작성과 초안 확인 단계에서는 Flash 모델로 빠르게 결과를 확인하고, 모든 것이 확정된 후 최종 생성에만 Multilingual v2를 사용하는 것입니다. 이 방식으로 크레딧 소모를 크게 줄이면서도 최종 결과물의 품질은 유지할 수 있습니다.

24 무료 플랜 100% 활용법 — 월 10,000자의 전략적 사용

24.1 무료 플랜의 현실적 가용 범위

무료 플랜에서 Multilingual 모델 기준 월 10,000자, Flash 모델 기준 월 20,000자가 제공됩니다. 10,000자는 대략 10분 분량의 음성에 해당합니다. 한 달에 10분이라는 양은 결코 많지 않지만, 전략적으로 사용하면 의미 있는 결과를 만들어낼 수 있습니다.

24.2 무료 플랜에서 가능한 것들

TTS를 사용한 음성 생성이 가능합니다. 월 10,000자(또는 Flash로 20,000자) 내에서 텍스트를 음성으로 변환할 수 있습니다. 보이스 라이브러리의 다양한 음성을 탐색하고 사용할 수 있습니다. 보이스 디자인으로 최대 3개의 커스텀 음성을 만들 수 있습니다. 효과음 생성과 음악 생성도 제한된 범위에서 가능합니다. 음성 인식(STT)도 소량 사용 가능합니다.

24.3 무료 플랜에서 불가능한 것들

음성 복제(인스턴트 클론, PVC)는 사용할 수 없습니다. 상업적 목적으로 생성한 오디오를 사용할 수 없습니다. 반드시 일레븐랩스를 출처로 표시해야 합니다. 한 번에 최대 2,500자까지만 입력할 수 있습니다.

24.4 무료 크레딧 극대화 전략

무료 플랜의 크레딧을 최대한 활용하기 위한 구체적인 전략을 제시합니다.

Flash 모델 우선 전략입니다. 무료 플랜에서 Flash 모델을 사용하면 20,000자를 사용할 수 있어, Multilingual의 두 배입니다. 최종 품질이 크게 중요하지 않은 개인 학습용이나 내부 참고용 음성에는 Flash 모델을 사용하여 사용 가능한 문자 수를 극대화합니다.

텍스트 압축 전략입니다. 같은 내용을 더 적은 문자로 전달할 수 있도록 텍스트를 간결하게 다듬습니다. 불필요한 수식어, 반복적인 표현, 장황한 설명을 줄이면 같은 정보를 더 적은 크레딧으로 음성화할 수 있습니다.

핵심 구간 선별 전략입니다. 모든 텍스트를 음성화하려 하지 않고, 가장 중요한 핵심 구간만 선별하여 TTS로 변환합니다. 예를 들어 블로그 글 전체를 음성화하기보다, 핵심 요약 부분만 변환하는 것입니다.

한 번에 완벽하게 전략입니다. 텍스트를 충분히 준비하고, 설정을 신중하게 조정한 후에 생성하여, 재생성 횟수를 최소화합니다.

24.5 무료 플랜의 장점 재발견

무료 플랜은 제한이 있지만, 일레븐랩스의 핵심 기술을 직접 경험하고 학습할 수 있는 훌륭한 출발점입니다. 다양한 음성을 탐색하고, TTS 텍스트 작성 기법을 연습하고, 오디오 태그 사용법을 익히는 등의 학습 과정에서는 무료 플랜으로 충분합니다. 이 과정에서 습득한 기술과 노하우는 나중에 유료 플랜으로 전환했을 때 크레딧을 훨씬 더 효율적으로 사용하는 데 도움이 됩니다.

25 초과 사용(Overage) 관리와 비용 예측

25.1 초과 사용의 구조

크리에이터 플랜 이상의 유료 플랜에서는 월간 기본 크레딧을 초과하여 사용할 수 있으며, 초과분에 대해 추가 비용이 부과됩니다. 초과 비용은 플랜이 높을수록 단위당 비용이 낮아지는 구조입니다.

TTS의 경우 크리에이터 플랜에서 초과 시 1,000자당 0.30달러, 프로 플랜에서 0.24달러, 스케일 플랜에서 0.18달러, 비즈니스 플랜에서 0.12달러가 부과됩니다.

대화형 AI의 경우 크리에이터 플랜에서 초과 시 분당 0.12달러, 스케일과 비즈니스 플랜에서 분당 0.096달러가 부과됩니다.

더빙의 경우 크리에이터 플랜에서 초과 시 분당 0.60달러, 비즈니스 플랜에서 분당 0.24달러가 부과됩니다.

25.2 초과 비용 관리 전략

초과 비용이 예상치 않게 늘어나는 것을 방지하기 위한 전략이 있습니다.

사용량을 정기적으로 모니터링합니다. 대시보드에서 현재 사용량을 확인하고, 월간 예산 내에서 관리합니다.

초과가 기본 크레딧의 30에서 50% 수준에 이르면 플랜 업그레이드를 고려합니다. 일레븐랩스의 가격 구조상, 초과 비용이 다음 단계 플랜의 가격에 접근하면 업그레이드하는 것이 더 경제적입니다.

프로젝트별로 크레딧 예산을 미리 계획합니다. 각 프로젝트에 필요한 대략적인 문자 수를 미리 산정하고, 월간 총 사용량이 기본 크레딧을 초과하지 않도록 계획합니다.

제5부. 실전 시나리오별 활용 가이드

26 유튜버와 영상 크리에이터를 위한 활용 시나리오

26.1 시나리오 1: 유튜브 영상 나레이션 제작

유튜브 영상에 전문적인 나레이션을 추가하는 시나리오입니다.

대본을 완성한 후, 보이스 라이브러리에서 영상의 톤과 주제에 맞는 음성을 선택합니다. 보이스오버 스튜디오에서 프로젝트를 생성하고, 대본 텍스트를 입력합니다. 짧은 구간을 먼저 Flash 모델로 테스트하여 음성과 설정이 적절한지 확인합니다. 설정이 확정되면 Multilingual v2 모델로 전체 대본을 변환합니다. 생성된 오디오를 다운로드하여 영상 편집 소프트웨어에서 영상과 결합합니다.

이 워크플로우를 사용하면, 전문 성우를 고용하지 않고도 일관된 품질의 나레이션을 제작할 수 있습니다. 특히 다작하는 유튜버에게 시간과 비용 면에서 큰 절감 효과가 있습니다.

26.2 시나리오 2: 다국어 자막 및 더빙으로 글로벌 도달 확대

한국어로 제작한 유튜브 영상을 영어, 일본어 등 다른 언어로 더빙하는 시나리오입니다.

완성된 영상 파일을 더빙 기능에 업로드하고, 대상 언어를 선택합니다. AI가 자동으로 원본 음성을 인식하고 번역하여, 원본 화자의 음성 특성을 유지하면서 대상 언어로 더빙합니다. 더빙 스튜디오에서 번역 텍스트를 검토하고 필요한 수정을 가합니다. 완성된 더빙 영상을 다운로드하여 유튜브에 다국어 버전으로 업로드합니다.

이를 통해 해외 시청자에게 도달할 수 있는 콘텐츠를 효율적으로 제작할 수 있습니다.

26.3 시나리오 3: 쇼츠/릴스용 효과적인 보이스오버

짧은 형식의 소셜 미디어 영상에 임팩트 있는 보이스오버를 추가하는 시나리오입니다.

짧은 대본(보통 200자에서 500자)을 작성합니다. 에너지 넘치거나 감성적인 톤의 음성을 선택하고, 안정성을 Creative로 설정하여 표현력을 극대화합니다. 짧은 분량이므로 크레딧 소모가 적어, 무료 플랜에서도 여러 개의 쇼츠 보이스오버를 제작할 수 있습니다.

27 팟캐스터와 오디오 콘텐츠 제작자를 위한 시나리오

27.1 시나리오 1: 팟캐스트 인트로/아웃트로 제작

팟캐스트의 인트로와 아웃투로를 전문적으로 제작하는 시나리오입니다.

인트로 대본을 작성합니다. "안녕하세요, [팟캐스트명]에 오신 걸 환영합니다. 오늘은 [주제]에 대해 이야기해보겠습니다."와 같은 짧은 텍스트입니다. 프로페셔널한 음성을 선택하고, Eleven Music으로 배경 음악을 생성합니다. TTS로 생성한 인트로 음성과 AI로 생성한 배경 음악을 오디오 편집 도구에서 결합하면, 전문 스튜디오에서 제작한 것과 같은 인트로가 완성됩니다.

한번 만든 인트로와 아웃투로는 모든 에피소드에서 반복 사용하므로, 크레딧 투자 대비 가치가 매우 높습니다.

27.2 시나리오 2: 대본 기반 팟캐스트 에피소드 제작

일인 팟캐스트를 AI 음성으로 제작하는 시나리오입니다.

에피소드 대본을 완성한 후 보이스오버 스튜디오에서 프로젝트를 생성합니다. 일관된 음성으로 전체 에피소드를 생성하고, 필요한 부분에 효과음을 추가합니다. 이 방식은 매주 에피소드를 발행해야 하는 부담감을 줄여주며, 성우 비용 없이 정기적인 콘텐츠 발행이 가능해집니다.

다만 무료 플랜에서는 상업적 사용이 금지되므로, 수익화를 목표로 하는 팟캐스트라면 유료 플랜이 필요합니다.

28.1 시나리오 1: 온라인 강의 나레이션 자동화

온라인 강의의 슬라이드 나레이션을 AI로 자동화하는 시나리오입니다.

각 슬라이드에 대한 설명 대본을 작성합니다. 보이스오버 스튜디오에서 슬라이드별로 섹션을 나누어 나레이션을 생성합니다. 교육 콘텐츠는 명확성이 가장 중요하므로, 안정성을 높게 설정하고 속도를 약간 느리게 조절합니다.

이 방식의 장점은, 강의 내용이 수정되었을 때 해당 부분의 나레이션만 재생성하면 된다는 것입니다. 전체를 다시 녹음할 필요 없이 특정 섹션만 수정할 수 있어 효율적입니다.

28.2 시나리오 2: 다국어 교육 자료 제작

하나의 교육 자료를 여러 언어로 제작하는 시나리오입니다.

한국어로 교육 콘텐츠의 나레이션을 완성한 후, 더빙 기능을 활용하여 영어, 일본어, 중국어 등 필요한 언어로 자동 번역 및 더빙합니다. 글로벌 기업의 내부 교육이나, 다국적 학생을 대상으로 하는 교육 기관에서 매우 유용합니다.

28.3 시나리오 3: 오디오 학습 자료 제작

교과서나 학습 자료를 오디오 형태로 변환하여, 학생들이 이동 중에도 학습할 수 있도록 하는 시나리오입니다.

학습 내용의 핵심을 정리한 텍스트를 TTS로 변환합니다. 각 챕터별로 구분하여 생성하고, 중요한 개념은 속도를 약간 느리게 설정하여 청취자가 이해하기 쉽게 합니다. 오디오 네이티브 기능을 학습 플랫폼에 임베드하면, 별도의 오디오 파일 관리 없이도 학습 자료의 오디오 버전을 제공할 수 있습니다.

29.1 시나리오 1: 제품 홍보 영상의 보이스오버

신제품 홍보 영상에 전문적인 보이스오버를 추가하는 시나리오입니다.

광고 카피를 작성하고, 브랜드 톤에 맞는 음성을 보이스 라이브러리에서 선택합니다. 안정성을 Creative로 설정하여 매력적이고 에너지 넘치는 톤을 만듭니다. 필요하다면 여러 버전을 생성하여 A/B 테스트를 진행할 수도 있습니다.

29.2 시나리오 2: IVR(자동응답) 시스템 음성 제작

고객 서비스 전화의 자동응답 시스템 안내 음성을 제작하는 시나리오입니다.

“안녕하세요, [회사명]에 전화해주셔서 감사합니다. 상담원 연결을 원하시면 1번을, 주문 조회를 원하시면 2번을 눌러주세요.”와 같은 안내 메시지를 TTS로 생성합니다. 안정성을 Robust로 설정하여 항상 일관된 톤의 안내 음성을 만듭니다.

29.3 시나리오 3: 프레젠테이션 오디오 가이드 제작

중요한 사업 프레젠테이션에 오디오 나레이션을 추가하는 시나리오입니다.

각 슬라이드의 발표자 노트를 정리하여 TTS로 변환합니다. 이를 프레젠테이션 파일에 삽입하면, 발표자 없이도 자동으로 진행되는 프레젠테이션을 만들 수 있습니다. 전시회 부스, 무인 안내 시스템, 온라인 배포용 프레젠테이션 등에 활용됩니다.

30 개발자와 기술 팀을 위한 API 활용 시나리오

30.1 시나리오 1: 앱에 TTS 통합

자체 애플리케이션에 일레븐랩스의 TTS 기능을 통합하는 시나리오입니다.

일레븐랩스는 RESTful API와 다양한 언어의 SDK를 제공합니다. API 키를 발급받고, 공식 문서에 따라 API 호출을 구현하면, 자체 앱에서 직접 텍스트를 음성으로 변환하는 기능을 제공할 수 있습니다.

뉴스 읽기 앱, 학습 앱, 접근성 도구, 게임 등 다양한 유형의 애플리케이션에서 활용 가능합니다.

30.2 시나리오 2: 대화형 AI 에이전트 개발

고객 서비스용 음성 AI 에이전트를 개발하는 시나리오입니다.

일레븐랩스의 대화형 AI 플랫폼을 활용하여, 실시간 음성 기반의 고객 서비스 에이전트를 구축합니다. 시스템 프롬프트에 에이전트의 역할, 응답 가이드라인, 지식 기반 등을 설정하고, Flash 모델을 선택하여 빠른 응답 속도를 보장합니다.

30.3 시나리오 3: 콘텐츠 파이프라인 자동화

대량의 콘텐츠를 자동으로 음성화하는 파이프라인을 구축하는 시나리오입니다.

API를 사용하여 텍스트 데이터베이스에서 콘텐츠를 읽어오고, 자동으로 TTS 변환을 수행하고, 생성된 오디오 파일을 지정된 저장소에 업로드하는 일련의 과정을 자동화합니다. 뉴스 미디어, 이러닝 플랫폼, 블로그 네트워크 등에서 대규모 오디오 콘텐츠를 효율적으로 생성할 수 있습니다.

31 작가와 출판 관계자를 위한 오디오북 시나리오

31.1 시나리오 1: 전자책을 오디오북으로 변환

기존에 출판된 전자책을 오디오북으로 변환하는 시나리오입니다.

전자책의 텍스트를 챕터별로 정리합니다. 보이스오버 스튜디오에서 챕터별로 프로젝트를 구성하고, 책의 분위기에 맞는 음성을 선택합니다. 발음 사전을 설정하여 등장인물의 이름이나 특수 용어의 발음을 통일합니다. 각 챕터를 순차적으로 생성하여, 완성된 오디오북을 만듭니다.

오디오북은 분량이 크므로 상당한 크레딧이 필요합니다. 프로 플랜 이상을 사용하는 것이 현실적이며, 프로젝트 계획 단계에서 총 문자 수를 미리 산정하여 필요한 크레딧을 확보하는 것이 중요합니다.

31.2 시나리오 2: 단편 소설의 오디오 버전 제작

짧은 단편 소설이나 에세이를 오디오 형태로 제작하여 팟캐스트나 소셜 미디어에서 공유하는 시나리오입니다.

짧은 분량이므로 크레딧 부담이 적고, 감정 표현이 풍부한 음성과 Creative 안정성 설정을 사용하면 문학적 감성이 살아있는 오디오를 만들 수 있습니다. Eleven v3의 오디오 태그를 활용하면 등장인물의 감정을 더욱 생생하게 표현할 수 있습니다.

제6부. 한계 인식과 현명한 대처

32 일레븐랩스의 정직한 한계 분석

32.1 기술적 한계

일레븐랩스의 활용법을 다루는 이 책에서, 그 한계를 정직하게 분석하는 것은 중요합니다. 한계를 정확히 알아야 현실적인 기대치를 설정하고 효과적인 활용 전략을 수립할 수 있기 때문입니다.

AI 음성은 아직 완벽하지 않습니다. 아무리 뛰어난 AI 모델이라도, 100% 완벽하게 사람의 음성을 재현하지는 못합니다. 특정 문맥에서 부자연스러운 억양이 나타나거나, 감정 표현이 과하거나 부족한 경우가 발생할 수 있습니다. 특히 매우 미묘한 감정의 뉘앙스나, 즉흥적인 대화에서의 자연스러운 반응은 아직 AI가 완벽하게 구현하기 어려운 영역입니다.

한국어 성능은 영어에 비해 약간의 차이가 있을 수 있습니다. 일레븐랩스의 AI 모델은 영어 데이터를 가장 풍부하게 학습했으므로, 영어 음성의 자연스러움이 가장 높습니다. 한국어를 포함한 다른 언어들도 높은 수준을 달성했지만, 미묘한 차이가 존재할 수 있습니다.

긴 문맥에서의 일관성에 제한이 있을 수 있습니다. 매우 긴 텍스트를 한꺼번에 변환할 때, 시작 부분과 끝 부분의 톤이나 스타일이 약간 달라질 수 있습니다. 보이스오버 스튜디오에서 섹션별로 나누어 생성하면 이 문제를 완화할 수 있습니다.

32.2 무료 플랜의 핵심 제한

무료 플랜의 가장 큰 제한은 상업적 사용 불가와 음성 복제 불가입니다. 상업적 목적의 콘텐츠 제작에는 반드시 유료 플랜이 필요하며, 자신만의 고유한 AI 음성을 만들려면 스타터 플랜 이상이 필요합니다.

32.3 한계에 대한 현명한 대처

기술적 한계에 대해서는 후처리를 통한 보완이 가능합니다. AI가 생성한 음성에서 부자연스러운 부분을 오디오 편집 도구로 수정하거나, 해당 구간만 재생성하는 등의 방법으로 최종 결과물의 품질을 높일 수 있습니다.

무료 플랜의 제한에 대해서는 용도를 정확히 구분하는 것이 중요합니다. 학습, 테스트, 개인 프로젝트에는 무료 플랜을 활용하고, 상업적 프로젝트에만 유료 플랜을 사용하는 방식으로 비용을 관리할 수 있습니다.

33.1 유료 전환을 고려해야 하는 신호

다음과 같은 신호가 나타나면 유료 플랜 전환을 진지하게 고려할 시점입니다.

매달 10,000자(또는 20,000자)의 무료 크레딧이 부족하여 원하는 작업을 완료하지 못하는 경우입니다. 생성한 음성을 상업적으로 사용해야 하는 프로젝트가 발생한 경우입니다. 자신의 목소리나 특정 음성을 복제하여 사용하고 싶은 경우입니다. 일레븐랩스가 업무의 핵심 도구로 자리 잡아 안정적인 사용이 필요한 경우입니다.

33.2 플랜 선택 가이드

스타터 플랜(월 5달러)은 개인 크리에이터가 소규모 프로젝트를 시작할 때 적합합니다. 인스턴트 음성 복제와 상업적 사용 권한이 필요하지만 사용량이 많지 않은 경우에 좋은 선택입니다.

크리에이터 플랜(월 11달러)은 정기적으로 콘텐츠를 제작하는 유튜버, 팟캐스터, 블로거 등에게 적합합니다. 월 100,000자는 대략 100분 분량의 음성에 해당하므로, 주 1회 정도의 장편 콘텐츠 제작이 가능합니다.

프로 플랜(월 99달러)은 전문적으로 오디오 콘텐츠를 제작하는 개인이나 소규모 팀에 적합합니다. 월 500,000자와 높은 동시성, 낮은 초과 비용이 안정적인 대량 생산을 가능하게 합니다.

스케일 플랜(월 330달러) 이상은 에이전시, 미디어 기업, 교육 기관 등 대규모 조직에 적합합니다. 팀 사용을 위한 다중 석과 대량의 크레딧이 제공됩니다.

33.3 연간 결제의 이점

모든 유료 플랜에서 연간 결제 옵션이 제공되며, 월간 결제 대비 상당한 할인이 적용됩니다. 장기적으로 일레븐랩스를 사용할 계획이라면, 연간 결제를 선택하는 것이 경제적입니다.

34.1 플랜별 상업적 이용 권한

무료 플랜에서는 생성한 오디오의 상업적 사용이 금지됩니다. 무료 플랜으로 생성한 오디오를 공개할 때는 반드시 일레븐랩스를 출처로 표시(attribution)해야 합니다. 상업적 사용에는 유료 콘텐츠 제작, 광고, 클라이언트 납품물, 판매용 제품 등이 포함됩니다.

스타터 플랜 이상의 모든 유료 플랜에서는 상업적 사용 권한이 부여됩니다. 생성한 오디오를 자유롭게 상업적 목적으로 활용할 수 있습니다.

34.2 보이스 라이브러리 음성의 라이선스

보이스 라이브러리에 공유된 커뮤니티 음성은 무료 상업적 사용 라이선스와 함께 제공되는 것으로 알려져 있습니다. 이는 유료 플랜 사용자가 해당 음성을 상업적 콘텐츠에 사용할 수 있다는 의미입니다. 다만 구체적인 라이선스 조건은 음성마다 다를 수 있으므로, 사용 전에 확인하는 것이 좋습니다.

아이코닉 마켓플레이스의 유명한 음성은 별도의 라이선스 조건이 적용됩니다. 사용 전 해당 음성의 라이선스 조건과 가격을 반드시 확인해야 합니다.

34.3 음성 복제와 윤리적 사용

음성 복제 기능은 강력한 만큼 윤리적 사용이 매우 중요합니다. 반드시 본인의 음성이거나, 복제 대상자의 명시적 동의를 받은 음성만 복제해야 합니다. 타인의 음성을 무단으로 복제하여 사용하는 것은 법적 책임을 수반할 수 있으며, 일레븐랩스의 이용약관 위반에 해당합니다.

생성된 AI 음성이 실제 사람의 발언으로 오인될 수 있는 상황에서는, AI로 생성되었음을 명시하는 것이 바람직합니다.

제7부. 부록

35 자주 묻는 질문(FAQ) 50선

가입과 계정 관련

질문 1. 일레븐랩스 가입에 신용카드가 필요한가요. 답변: 무료 플랜은 신용카드 없이 이메일이나 구글 계정으로 가입할 수 있습니다.

질문 2. 가입 후 바로 사용할 수 있나요. 답변: 네, 가입 즉시 월 10,000자의 크레딧이 제공되며 모든 무료 기능을 바로 이용할 수 있습니다.

질문 3. 모바일에서도 사용할 수 있나요. 답변: 네, 웹 브라우저와 모바일 앱(iOS, Android)을 통해 사용 가능합니다.

질문 4. 계정 하나로 여러 기기에서 사용할 수 있나요. 답변: 네, 동일 계정으로 여러 기기에서 접속 가능하며, 크레딧은 계정 단위로 관리됩니다.

질문 5. 계정을 삭제하고 싶으면 어떻게 하나요. 답변: 설정 메뉴에서 계정 삭제 옵션을 찾거나, 고객 지원에 문의하여 삭제를 요청할 수 있습니다.

크레딧 관련

질문 6. 무료 플랜에서 한 달에 몇 자를 사용할 수 있나요. 답변: Multilingual 모델 기준 월 10,000자, Flash 모델 기준 월 20,000자입니다.

질문 7. 크레딧은 언제 갱신되나요. 답변: 가입일 또는 결제일로부터 30일(무료 플랜) 또는 결제 주기마다 갱신됩니다.

질문 8. 남은 크레딧은 다음 달로 이월되나요. 답변: 아닙니다. 기본 제공 크레딧은 이월되지 않습니다.

질문 9. 크레딧 잔액은 어디서 확인하나요. 답변: 대시보드의 프로필 영역이나 설정 메뉴에서 확인할 수 있습니다.

질문 10. 한 번에 최대 몇 자까지 입력할 수 있나요. 답변: 웹사이트 음성 합성에서 무료 플랜은 2,500자, 유료 플랜은 5,000자입니다.

TTS 관련

질문 11. 한국어 TTS가 가능한가요. 답변: 네, Multilingual 모델에서 한국어를 포함한 29개 이상의 언어를 지원합니다.

질문 12. Flash 모델과 Multilingual 모델의 차이는 무엇인가요. 답변: Flash는 빠른 속도와 낮은 크레딧 소모가 장점이고, Multilingual은 높은 품질과 폭넓은 언어 지원이 장점입니다.

질문 13. 생성된 음성을 다운로드할 수 있나요. 답변: 네, MP3 등의 형식으로 다운로드할 수 있습니다.

질문 14. Eleven v3 모델은 무엇인가요. 답변: 오디오 태그를 통한 감정 제어, 효과음 삽입 등의 고급 기능을 지원하는 최신 모델입니다.

질문 15. TTS 결과가 마음에 들지 않으면 어떻게 하나요. 답변: 텍스트를 수정하거나, 음성 설정(안정성, 유사도 등)을 조정하거나, 다른 음성을 선택하여 재생성할 수 있습니다. 재생성 시 크레딧이 추가로 소모됩니다.

음성 복제 관련

질문 16. 무료 플랜에서 음성 복제가 가능한가요. 답변: 아닙니다. 인스턴트 클론은 스타터 플랜 이상, 프로페셔널 클론은 크리에이터 플랜 이상에서 사용 가능합니다.

질문 17. 다른 사람의 목소리를 복제해도 되나요. 답변: 해당 사람의 명시적 동의가 있어야만 복제할 수 있습니다. 무단 복제는 이용약관 위반이며 법적 문제를 야기할 수 있습니다.

질문 18. 인스턴트 클론에 필요한 음성 샘플 길이는 얼마인가요. 답변: 최소 약 1분 이상이 권장됩니다. 더 긴 샘플일수록 좋은 결과를 얻을 수 있습니다.

질문 19. 프로페셔널 클론(PVC)에 필요한 샘플은 얼마인가요. 답변: 일반적으로 30분 이상의 고품질 샘플이 권장됩니다.

질문 20. 복제된 음성의 품질이 기대에 미치지 못할 때는 어떻게 하나요. 답변: 더 깨끗하고 다양한 감정이 포함된 샘플을 사용하여 다시 시도할 수 있습니다. Voice Isolator로 샘플의 배경 소음을 제거한 후 사용하면 품질이 향상될 수 있습니다.

더빙 관련

질문 21. 더빙은 어떤 언어를 지원하나요. 답변: 29개 이상의 언어를 지원하며, 한국어도 포함됩니다.

질문 22. 더빙 결과의 번역을 수정할 수 있나요. 답변: 더빙 스튜디오에서 번역된 텍스트를 직접 수정할 수 있습니다.

질문 23. 더빙에 크레딧이 얼마나 소모되나요. 답변: 분 단위로 소모되며, 소모량은 콘텐츠의 길이와 복잡도에 따라 달라집니다.

효과음과 음악 관련

질문 24. 효과음 생성은 무료로 사용할 수 있나요. 답변: 무료 플랜에서도 제한된 범위에서 사용 가능하지만, 생성 횟수에 제한이 있습니다.

질문 25. Eleven Music으로 생성한 음악을 상업적으로 사용할 수 있나요. 답변: 유료 플랜에서 생성한 음악은 상업적 사용이 가능한 로열티 프리 음악입니다.

질문 26. 효과음의 길이를 지정할 수 있나요. 답변: 프롬프트에 원하는 길이를 명시할 수 있지만, 정확한 길이 제어는 제한적일 수 있습니다.

음성 설정 관련

질문 27. 안정성(Stability) 설정을 어떻게 조정해야 하나요. 답변: 일관된 톤이 필요하면 높게, 표현력이 필요하면 낮게 설정합니다. 용도에 따라 실험하면서 최적값을 찾으세요.

질문 28. 속도를 조절할 수 있나요. 답변: 네, 0.7(느리게)에서 1.2(빠르게)까지 조절 가능합니다. 극단적인 값은 품질에 영향을 줄 수 있습니다.

질문 29. 오디오 태그는 어떤 모델에서 사용할 수 있나요. 답변: 오디오 태그는 Eleven v3 모델에서 사용할 수 있습니다.

질문 30. SSML 태그를 사용할 수 있나요. 답변: 일부 모델에서 break 태그나 phoneme 태그 등의 SSML을 지원합니다. 다만 Eleven v3에서는 SSML break 태그가 지원되지 않으며, 오디오 태그와 구두점으로 대체합니다.

보이스오버 스튜디오 관련

질문 31. 보이스오버 스튜디오에서 프로젝트를 만들면 크레딧이 소모되나요. 답변: 프로젝트 생성과 텍스트 입력 자체는 크레딧을 소모하지 않습니다. 실제 음성 생성(Convert) 시에만 소모됩니다.

질문 32. 하나의 프로젝트에서 여러 음성을 사용할 수 있나요. 답변: 네, 텍스트의 각 부분에 서로 다른 음성을 지정할 수 있습니다.

질문 33. 무료 플랜에서 프로젝트를 몇 개까지 만들 수 있나요. 답변: 최대 3개의 프로젝트를 만들 수 있습니다.

보이스 라이브러리 관련

질문 34. 보이스 라이브러리의 음성은 무료인가요. 답변: 네, 보이스 라이브러리의 커뮤니티 음성은 무료로 탐색하고 사용할 수 있습니다.

질문 35. 보이스 라이브러리에서 한국어 음성을 찾을 수 있나요. 답변: 네, 언어 필터를 한국어로 설정하면 한국어에 적합한 음성을 찾을 수 있습니다.

질문 36. 내가 만든 음성을 보이스 라이브러리에 공유할 수 있나요. 답변: 네, 프로페셔널

음성 복제(PVC)로 만든 음성을 커뮤니티에 공유하고, 다른 사용자가 사용할 때 보상을 받을 수 있습니다.

대화형 AI 관련

질문 37. 무료 플랜에서 대화형 AI를 사용할 수 있나요. 답변: 제한된 범위에서 사용 가능합니다. 본격적인 서비스 운영에는 유료 플랜이 필요합니다.

질문 38. 대화형 AI 에이전트를 만들려면 프로그래밍 지식이 필요한가요. 답변: 기본적인 에이전트는 웹 인터페이스에서 설정할 수 있지만, 고급 통합에는 API 사용 경험이 도움이 됩니다.

오디오 네이티브 관련

질문 39. 오디오 네이티브를 내 블로그에 설치할 수 있나요. 답변: 네, 간단한 임베드 코드를 웹사이트에 삽입하면 됩니다.

질문 40. 오디오 네이티브는 어떤 플랜에서 사용할 수 있나요. 답변: 플랜별 사용 가능 여부와 조건은 일레븐랩스 공식 사이트에서 확인하시기 바랍니다.

상업적 이용 관련

질문 41. 무료 플랜에서 생성한 음성을 유튜브에 올릴 수 있나요. 답변: 올릴 수는 있지만, 상업적 사용은 금지되므로 수익화(광고 수익 등)는 할 수 없으며 일레븐랩스 출처 표시가 필요합니다.

질문 42. 유료 플랜에서 생성한 음성의 저작권은 누구에게 있나요. 답변: 구체적인 저작권 조건은 일레븐랩스의 이용약관을 확인하시기 바랍니다.

질문 43. 생성한 음성을 재판매할 수 있나요. 답변: 이용약관에 따라 다를 수 있으므로, 구체적인 재판매 조건은 공식 약관을 확인해야 합니다.

기술 관련

질문 44. API를 사용하려면 어떤 플랜이 필요한가요. 답변: API는 무료 플랜을 포함한 모든 플랜에서 사용 가능합니다.

질문 45. 생성된 오디오의 파일 형식은 무엇인가요. 답변: MP3가 기본이며, API를 통해 다른 형식을 지정할 수도 있습니다.

질문 46. 오프라인에서 사용할 수 있나요. 답변: 아닙니다. 일레븐랩스는 클라우드 기반 서비스이므로 인터넷 연결이 필요합니다.

질문 47. 일레븐랩스에 입력한 데이터는 안전한가요. 답변: 데이터 보안에 대한 구체적인 정책은 개인정보처리방침을 확인하시기 바랍니다. 민감한 정보 입력 시 주의가 필요합니다.

질문 48. 기능이 자주 업데이트되나요. 답변: 네, 일레븐랩스는 지속적으로 새로운 기능과

모델을 추가하고 있습니다.

질문 49. 기술적 문제가 있으면 어디에 문의하나요. 답변: 헬프센터(help.elevenlabs.io)나 고객 지원 채널을 통해 문의할 수 있습니다.

질문 50. 일레븐랩스와 유사한 서비스가 있나요. 답변: AI TTS 분야에는 여러 서비스가 존재합니다. 각 서비스마다 강점과 특성이 다르므로, 자신의 용도에 가장 적합한 서비스를 선택하는 것이 중요합니다.

36 기능별 크레딧 소모량 참고표

아래 정보는 공개된 자료와 사용자 경험을 종합한 대략적인 참고값입니다. 실제 소모량은 사용 조건에 따라 달라질 수 있으므로, 정확한 정보는 일레븐랩스 공식 문서를 참고하시기 바랍니다.

TTS 크레딧 소모 (문자 단위)

TTS(Multilingual v2): 입력 문자 수 = 크레딧 소모량. 1,000자 입력 시 1,000자 크레딧 소모. 약 1분 분량의 음성에 해당합니다.

TTS(Flash): 입력 문자 수 = 크레딧 소모량이지만, 같은 플랜에서 Multilingual 대비 2배의 문자 한도가 적용됩니다.

더빙 크레딧 소모 (분 단위)

크리에이터 플랜: 월 50분 포함, 초과 시 분당 약 0.60달러. 프로 플랜: 월 250분 포함, 초과 시 분당 약 0.40달러 내외. 스케일 플랜: 월 1,000분 포함. 비즈니스 플랜: 월 5,500분 포함.

음성 변환 및 음성 분리 크레딧 소모 (분 단위)

크리에이터 플랜: 월 100분 포함. 프로 플랜: 월 500분 포함. 스케일 플랜: 월 2,000분 포함. 비즈니스 플랜: 월 11,000분 포함.

음성 인식(STT) 크레딧 소모 (시간 단위)

API: 크리에이터 약 63시간, 프로 약 300시간, 스케일 약 1,100시간. UI: 크리에이터 약 5시간, 프로 약 25시간, 스케일 약 94시간.

효과음 크레딧 소모 (생성 횟수 단위)

크리에이터 플랜: 월 500회. 프로 플랜: 월 2,500회. 스케일 플랜: 월 10,000회. 비즈니스 플랜: 월 55,000회.

대화형 AI 크레딧 소모 (분 단위)

크리에이터 플랜: 월 250분 포함. 프로 플랜: 월 1,100분 포함. 스케일 플랜: 월 3,600분 포함. 비즈니스 플랜: 월 13,750분 포함.

이 참고표를 활용하여, 프로젝트에 필요한 대략적인 크레딧 소모를 미리 예측하고 적절한 플랜을 선택할 수 있습니다.

37 용어 사전

Audio Native(오디오 네이티브): 웹페이지에 임베드할 수 있는 오디오 플레이어로, 웹 콘텐츠를 자동으로 음성으로 변환하여 재생합니다.

CMU Arpabet: 카네기 멜론 대학교에서 개발한 음성 기호 체계로, 영어 발음을 ASCII 문자로 표기합니다. 일레븐랩스의 발음 사전에서 사용됩니다.

Conversational AI(대화형 AI): 실시간 음성 기반의 AI 에이전트를 구축할 수 있는 플랫폼입니다. 고객 서비스, 교육, 인터랙티브 콘텐츠 등에 활용됩니다.

Dubbing(더빙): 오디오나 비디오 콘텐츠를 다른 언어로 자동 번역하고, 원본 화자의 음성 특성을 유지하면서 음성을 합성하는 기능입니다.

Dubbing Studio(더빙 스튜디오): 더빙 결과를 세부적으로 편집할 수 있는 고급 인터페이스입니다.

Eleven Music: 자연어 프롬프트로 스튜디오급 음악을 생성하는 AI 음악 생성 도구입니다.

Eleven v3: 오디오 태그를 통한 감정 제어, 효과음 삽입 등의 고급 기능을 지원하는 최신 TTS 모델입니다.

Flash 모델: 속도에 최적화된 TTS 모델로, 낮은 지연 시간과 효율적인 크레딧 소모가 특징입니다.

Iconic Marketplace(아이코닉 마켓플레이스): 유명인의 공인된 AI 음성을 라이선스하여 사용할 수 있는 프리미엄 마켓플레이스입니다.

Instant Clone(인스턴트 클론): 짧은 음성 샘플로 빠르게 음성을 복제하는 기능입니다. 스타터 플랜 이상에서 사용 가능합니다.

IPA(International Phonetic Alphabet, 국제 음성 기호): 전 세계 언어의 발음을 표기하는 국제 표준 기호 체계입니다.

kbps(kilobits per second): 오디오 품질 단위로, 숫자가 높을수록 고품질입니다. 일레븐랩스에서 무료/스타터 플랜은 128kbps, 크리에이터 이상은 192kbps를 제공합니다.

Multilingual v2: 29개 이상의 언어를 지원하는 고품질 TTS 모델로, 풍부한 감정 표현과 자연스러운 억양이 특징입니다.

Overage(초과 사용): 플랜의 기본 크레딧을 초과하여 사용하는 것으로, 추가 비용이 발생합니다.

PLS(Pronunciation Lexicon Specification): 발음 사전을 정의하는 W3C 표준 파일 형식입니다.

Professional Voice Clone(PVC, 프로페셔널 음성 복제): 긴 음성 샘플을 사용하여 높은 정확도로 음성을 복제하는 기능입니다. 크리에이터 플랜 이상에서 사용 가능합니다.

Pronunciation Dictionary(발음 사전): 특정 단어나 구문의 발음을 미리 지정하는 파일로, 보이스오버 스튜디오와 더빙 스튜디오에서 사용할 수 있습니다.

Scribe: 일레븐랩스의 자체 음성 인식(STT) 모델로, 90개 이상의 언어에 대한 고정밀 전사를 제공합니다.

Sound Effects(효과음): 텍스트 프롬프트를 입력하면 AI가 해당하는 효과음을 생성하는 기능입니다.

Speech-to-Text(STT, 음성 인식): 음성을 텍스트로 변환하는 기술입니다.

SSML(Speech Synthesis Markup Language): 음성 합성을 제어하기 위한 마크업 언어로, 일부 일레븐랩스 모델에서 지원됩니다.

Stability(안정성): TTS 생성 시 음성의 일관성과 변동성을 조절하는 파라미터입니다.

Text-to-Speech(TTS, 텍스트 투 스피치): 텍스트를 음성으로 변환하는 기술입니다.

Turn-Taking(턴테이킹): 대화형 AI에서 AI와 사용자가 자연스럽게 발언을 주고받는 기술입니다.

Voice Changer(음성 변환): 녹음된 음성의 목소리 특성을 다른 음성으로 변환하는 기능입니다.

Voice Design(보이스 디자인): 텍스트 설명을 바탕으로 완전히 새로운 합성 음성을 생성하는 기능입니다.

Voice Isolator(음성 분리): AI를 사용하여 오디오에서 음성만 추출하고 배경 소음을 제거하는 기능입니다.

Voice Library(보이스 라이브러리): 커뮤니티가 공유한 다양한 음성을 탐색하고 사용할 수 있는 마켓플레이스입니다.

Voiceover Studio(보이스오버 스튜디오): 타임라인 기반의 장편 오디오 콘텐츠 제작 환경입니다.

Watermark(워터마크): 무료 플랜에서 생성된 일부 콘텐츠에 적용될 수 있는 식별 표시입니다.

마치며

이 책에서 우리는 일레븐랩스의 모든 핵심 기능을 살펴보았습니다. TTS의 기본 원리부터 음성 복제, 더빙, 효과음, AI 음악, 대화형 AI에 이르는 기능별 활용 전략을 수립하고, 오디오 태그와 텍스트 작성 기법을 통한 품질 극대화 방법을 익혔으며, 크레딧 관리와 비용 최적화 전략을 다루고, 다양한 직업군과 상황에 맞는 실전 시나리오를 제시했습니다.

다시 한번 강조하고 싶은 핵심 메시지가 있습니다. AI 음성 기술은 "사람을 대체하는 것"이 아니라 "사람의 가능성을 확장하는 것"입니다. 전문 성우를 고용할 예산이 없어 음성 콘텐츠를 포기해야 했던 개인 크리에이터가, 이제는 전문적인 품질의 나레이션을 스스로 만들 수 있습니다. 한 가지 언어로만 소통할 수 있었던 콘텐츠가, 이제는 29개 이상의 언어로 전 세계에 전달될 수 있습니다. 매년 새로 녹음해야 했던 교육 자료가, 이제는 텍스트 수정만으로 즉시 업데이트될 수 있습니다.

일레븐랩스의 기술은 빠르게 발전하고 있습니다. 이 책에서 다룬 정보 중 일부는 시간이 지남에 따라 변경될 수 있습니다. 새로운 모델이 출시되고, 기존 기능이 개선되며, 요금제가 조정될 수 있습니다. 최신 정보는 항상 일레븐랩스 공식 웹사이트(elevenlabs.io)와 공식 문서(elevenlabs.io/docs)에서 확인하시기 바랍니다.

마지막으로, 기술은 도구이며 그 가치는 사용하는 사람의 창의성과 목적에 의해 결정됩니다. 일레븐랩스라는 강력한 도구를 여러분의 아이디어, 전문성, 그리고 열정과 결합하여 새로운 가치를 만들어내시기를 응원합니다.

이 책을 읽어주셔서 감사합니다.

면책 조항

이 책에 포함된 정보는 2026년 3월 기준으로 공개적으로 확인 가능한 자료를 바탕으로 작성되었습니다. 일레븐랩스의 기능, 요금제, 크레딧 소모량, 지원 언어 등은 서비스 제공자의 정책 변경에 따라 달라질 수 있으며, 이 책의 내용이 최신 상태를 반영하지 못할 수 있습니다. 정확한 최신 정보는 일레븐랩스 공식 웹사이트(elevenlabs.io)와 헬프센터(help.elevenlabs.io)에서 직접 확인하시기 바랍니다.

이 책에서 언급된 크레딧 소모량과 기능별 세부 사항은 공개된 자료와 사용자 경험을 종합한 대략적인 참고값이며, 공식적으로 보장된 수치가 아닙니다. 실제 소모량과 기능 동작은 사용 조건에 따라 달라질 수 있습니다.