

# Mejora #1: Sistema de Conversaciones de Voz en Tiempo Real

---

## Descripción General





---

Implementación completa del **Sistema de Conversaciones de Voz en Tiempo Real con Análisis Profundo**, la primera mejora propuesta en el análisis de IA para SpeaklyPlan.








## Características Implementadas

---





### 1. Conversación por Voz Continua

-  Reconocimiento de voz continuo (sin necesidad de presionar botones repetidamente)
-  Detección automática cuando el usuario termina de hablar
-  Transcripción en tiempo real con feedback visual
-  Manejo de interrupciones y reinicio automático





### 2. Análisis Fonético Profundo con IA

-  Evaluación de cada fonema pronunciado
-  Identificación de problemas específicos (th, r, v, etc.)
-  Puntuación de pronunciación (0-100)
-  Puntuación de fluidez (0-100)
-  Análisis de acento y similitud con hablante nativo (0-100)
-  Detección de patrones de errores recurrentes
-  Generación automática de ejercicios personalizados

### 3. Corrección en Tiempo Real Suave

-  El AI responde naturalmente incorporando correcciones
-  Feedback constructivo sin interrumpir el flujo
-  Sugerencias específicas para sonidos problemáticos
-  Toast notifications con tips de pronunciación

### 4. Simulación de Acentos Reales

-  Soporte para múltiples acentos: American, British, Indian, Australian
-  Características específicas de cada acento
-  Prompts de práctica adaptados a cada acento
-  Síntesis de voz configurada según el acento objetivo

## Arquitectura Técnica

---

### Servicios Creados

#### 1. voice-conversation-service.ts

Servicio principal de análisis de voz:

- `analyzeVoicePronunciation()` : Análisis profundo de pronunciación
- `generateVoiceResponse()` : Generación de respuestas conversacionales

- `detectPronunciationPatterns()` : Detección de patrones de error
- `generateCustomExercises()` : Creación de ejercicios personalizados

## 2. API Endpoints

### `/api/tutor/voice/analyze` (POST)

- Analiza el transcript de voz del usuario
- Devuelve: `pronunciationScore`, `fluencyScore`, `accentScore`, `phonemeErrors`, `suggestions`
- Guarda la sesión en la base de datos
- Detecta patrones de errores recurrentes

### `/api/tutor/voice/analyze` (GET)

- Obtiene historial de sesiones de voz
- Calcula estadísticas agregadas
- Detecta mejoras a lo largo del tiempo

### `/api/tutor/voice/stream` (POST)

- Genera respuesta conversacional del tutor
- Determina si debe corregir errores críticos
- Mantiene el flujo natural de conversación

## Componentes Creados

### 1. `voice-conversation.tsx`

Componente principal de conversación de voz:

- Manejo de Web Speech API
- Reconocimiento continuo de voz
- Síntesis de voz con acentos
- Análisis en tiempo real
- Visualización de estadísticas de sesión
- Historial de conversación

### 2. `voice-client.tsx`

Cliente de página con:

- Selector de acento objetivo
- Información de uso
- Grid de características
- Integración con layout principal

### 3. `page.tsx`

Página de voz con:

- Autenticación requerida
- Suspense loading
- Integración con navegación

## Base de Datos

### Nuevo Modelo: VoiceSession

```
model VoiceSession {
  id                String
  userId            String
  conversationId    String?
  transcript        String
  pronunciationScore Float
  fluencyScore      Float
  accentScore       Float
  phonemeErrors     Json?
  suggestions       String[]
  createdAt         DateTime
}
```

#### Relaciones:

- User → voiceSessions (one-to-many)
- ChatConversation → voiceSessions (one-to-many)

## Interfaz de Usuario

### Vista Principal

#### 1. Indicador de Estado

- Círculo animado grande (Listening/Speaking/Ready)
- Estado visual claro con colores y animaciones
- Iconos contextuales

#### 2. Transcripción en Vivo

- Muestra lo que el usuario está diciendo en tiempo real
- Diseño limpio con fondo azul

#### 3. Control de Conversación

- Botón grande “Start Speaking” / “Stop Listening”
- Deshabilitado durante análisis
- Cambio de color según estado

#### 4. Estadísticas de Sesión

- Pronunciation score con barra de progreso
- Fluency score con barra de progreso
- Accent similarity con barra de progreso

#### 5. Feedback del Análisis

- Fortalezas (verde)
- Sugerencias (azul)
- Sonidos a practicar (amarillo)

#### 6. Historial de Conversación

- Últimos 6 intercambios
- Muestra puntuaciones por turno
- Diseño diferenciado para usuario vs tutor

## Selector de Acentos

- Grid 2x2 en móvil, 4 columnas en desktop
- Muestra características principales de cada acento
- Selección visual clara

## Banner Promocional

- Visible en sidebar del tutor (desktop)
- Visible en menú móvil
- Badge “NEW” destacado
- Gradiente azul a púrpura
- Ícono de radio/voz



## Métricas y Analytics

---

### Por Sesión

- Pronunciation Score (0-100)
- Fluency Score (0-100)
- Accent Score (0-100)
- Phoneme Errors detectados
- Suggestions generadas

### Históricas

- Total de sesiones
- Promedio de pronunciación
- Promedio de fluidez
- Promedio de acento
- Mejora reciente (%)

### Patrones

- Fonemas problemáticos recurrentes
- Ejercicios personalizados generados
- Tracking de mejora por fonema



## Uso

---

### Para el Usuario

#### 1. Acceder a Voice Practice

- Desde el Tutor AI, hacer clic en el banner “Voice Practice AI”
- O navegar a `/tutor/voice`

#### 2. Seleccionar Acento

- Elegir entre American, British, Indian, Australian

#### 3. Iniciar Conversación

- Clic en “Start Speaking”
- Hablar naturalmente en inglés

- El AI transcribe y analiza en tiempo real
- El tutor responde con voz

#### 4. **Ver Feedback**

- Observar puntuaciones en tiempo real
- Leer sugerencias específicas
- Ver fonemas a practicar

#### 5. **Mejorar Continuamente**

- El sistema detecta patrones
- Genera ejercicios personalizados
- Trackea mejora a lo largo del tiempo

## **Tecnologías Utilizadas**

---

### **Frontend**

- Web Speech API (Recognition & Synthesis)
- React Hooks (useState, useRef, useEffect)
- Lucide Icons
- Shadcn/ui Components
- Toast notifications

### **Backend**

- Next.js API Routes
- Prisma ORM
- PostgreSQL
- Abacus AI LLM API (gpt-4o-mini)






### **IA**

- Análisis fonético con LLM
- Detección de patrones con ML
- Generación de ejercicios personalizados
- Respuestas conversacionales contextuales





## **Impacto Esperado**



---

### **Métricas de Éxito**

-  Pronunciación evaluada en cada sesión
-  Feedback inmediato y personalizado
-  Detección de patrones de error
-  Ejercicios adaptativos generados
-  Múltiples acentos soportados

### **Diferenciadores vs Competencia**

-  Conversación continua (no press-to-talk)
-  Análisis fonético profundo con IA
-  Detección de patrones personalizados
-  Ejercicios generados automáticamente

-  Múltiples acentos nativos
-  Feedback en tiempo real sin interrumpir flujo

## Próximos Pasos Sugeridos

---

### Fase 1: Pulido (Opcional)

1. Agregar más acentos (Canadiense, Irlandés, etc.)
2. Mejorar detección de fonemas específicos
3. Agregar visualización de ondas de voz
4. Implementar comparación lado-a-lado con nativo

### Fase 2: Expansión

1. Integrar con escenarios profesionales (Mejora #2)
2. Agregar práctica de vocabulario por voz
3. Implementar simulaciones de meetings con voz
4. Crear challenges de pronunciación

### Fase 3: Gamificación

1. Logros por mejora en fonemas específicos
2. Challenges de acento
3. Leaderboards de pronunciación
4. Badges por sesiones de voz

## Notas de Implementación

---

### Limitaciones Conocidas

- Web Speech API solo funciona en Chrome/Edge (no Firefox/Safari mobile)
- Requiere permisos de micrófono del navegador
- Requiere conexión a internet estable
- Análisis consume tokens de LLM (considerar optimización)

### Consideraciones de UX

- Toast notifications no intrusivas
- Indicadores visuales claros de estado
- Feedback inmediato pero no abrumador
- Diseño responsive para móvil

### Seguridad

- Transcripts almacenados en base de datos propia
- No se guardan archivos de audio (solo transcripts)
- Análisis procesado server-side
- Protección de rutas con autenticación




## Conclusión

---

La Mejora #1 está completamente implementada y lista para uso. Proporciona una experiencia de práctica de voz revolucionaria que diferencia a SpeaklyPlan de cualquier otra plataforma de aprendizaje de inglés en el mercado.

La combinación de reconocimiento continuo, análisis profundo con IA, y feedback personalizado crea una experiencia de aprendizaje única que prepara a los usuarios para conversaciones reales en contextos profesionales.

---

**Estado:**  Implementado completamente

**Fecha:** Octubre 2025

**Versión:** 1.0.0