Sage Voice Analysis Architecture

This document provides a comprehensive overview of the Sage voice analysis system, featuring a hybrid client-server architecture for research-grade vocal biomarker analysis. It's designed to help contributors understand the complete system design from user interaction to clinical insights.

System Overview

Sage is a research-grade vocal analysis platform that combines immediate local analysis (iOS SFVoiceAnalytics) with comprehensive cloud analysis (Parselmouth/Praat) to provide clinical-quality voice biomarkers including F0, jitter, shimmer, and HNR measurements.

High-Level System Architecture

```
flowchart TD
    subgraph "iOS App"
        A1[User Interface]
        A2[Local Analysis]
        A3[Authentication]
    end

    subgraph "Cloud Infrastructure"
        B1[Firebase Storage]
        B2[Cloud Functions]
        B3[Firestore Database]
    end

    subgraph "Analysis Pipeline"
        C1[SFVoiceAnalytics<br/>Local Analysis]
        C2[Parselmouth<br/>Cloud Analysis]
        C3[Clinical Assessment]
    end

    A1 --> A2
    A2 --> C1
    A1 --> B1
    B1 --> B2
    B2 --> C2
    C2 --> C3
    C3 --> B3
    B3 --> A1

    classDef ios fill:#e3f2fd,stroke:#1976d2,stroke-width:2px
    classDef cloud fill:#e8f5e8,stroke:#388e3c,stroke-width:2px
    classDef analysis fill:#fff3e0,stroke:#f57c00,stroke-width:2px

    class A1,A2,A3 ios
    class B1,B2,B3 cloud
    class C1,C2,C3 analysis
```

Domain-Driven Architecture Layers

Application Structure

```
graph TD
    subgraph "App Layer"
        A1[SageApp.swift<br/>Entry Point]
        A2[ContentView.swift<br/>Tab Navigation]
    end

    subgraph "Features Layer"
        B1[OnboardingJourneyView<br/>Voice Setup with Baseline]
        B2[SessionsView<br/>Daily Recording]
        B3[HomeView<br/>Today's Voice Analysis]
        B4[VoiceDashboardView<br/>Longitudinal Trends]
        B5[AuthChoiceView<br/>Authentication]
    end

    subgraph "ViewModels"
        C1[OnboardingJourneyViewModel<br/>Recording & Baseline Logic]
        C2[AuthViewModel<br/>Authentication State]
        C3[SessionsViewModel<br/>Daily Sessions]
    end

    subgraph "Domain Services"
        D1[HybridVocalAnalysisService<br/>Analysis Orchestration]
        D2[VocalBaselineService<br/>Baseline Management]
        D3[LocalVoiceAnalyzer<br/>iOS SFVoiceAnalytics]
    end

    subgraph "Domain Models"
        E1[VocalBiomarkers<br/>Clinical Models]
        E2[VocalBaseline<br/>User Baseline]
        E3[UserProfile<br/>User Data]
    end

    subgraph "Infrastructure Services"
        F1[AuthService<br/>Firebase Auth]
        F2[MicrophonePermissionManager<br/>Audio Permissions]
        F3[UserProfileRepository<br/>Profile Storage]
        F4[VoiceAnalysisRepository<br/>Analysis Storage]
    end

    A1 --> A2
    A2 --> B1
    A2 --> B2
    A2 --> B3
    A2 --> B4
    A2 --> B5
    B1 --> C1
    B5 --> C2
```

```
B2 --> C3
C1 --> D1
C1 --> D2
D1 --> D3
D1 --> E1
D2 --> E2
D2 --> E3
C1 --> F1
C1 --> F2
C1 --> F3
D1 --> F4

classDef app fill:#ffebee,stroke:#d32f2f,stroke-width:2px
classDef feature fill:#e3f2fd,stroke:#1976d2,stroke-width:1px
classDef viewmodel fill:#f3e5f5,stroke:#7b1fa2,stroke-width:1px
classDef domain fill:#e8f5e8,stroke:#388e3c,stroke-width:1px
classDef model fill:#fff3e0,stroke:#f57c00,stroke-width:1px
classDef infrastructure fill:#fce4ec,stroke:#c2185b,stroke-width:1px

class A1,A2 app
class B1,B2,B3,B4,B5 feature
class C1,C2,C3 viewmodel
class D1,D2,D3 domain
class E1,E2,E3 model
class F1,F2,F3,F4 infrastructure
```

Core User Flows

Hybrid Analysis Pipeline

```
sequenceDiagram
    participant U as User
    participant S as SessionsView
    participant H as HybridVocalAnalysisService
    participant L as LocalVoiceAnalyzer
    participant C as CloudVoiceAnalysisService
    participant F as Firestore

    U->>S: Tap Record (5s)
    S->>H: analyzeVoice(recording)

    Note over H: Phase 1: Immediate Local Analysis
    H->>L: analyzeImmediate(audioURL)
    L->>L: SFVoiceAnalytics<br/>F0 extraction only
    L->>H: BasicVoiceMetrics
    H->>S: Immediate Results<br/>(< 5 seconds)

    Note over H: Phase 2: Comprehensive Cloud Analysis
    H->>C: uploadAndAnalyze(recording)
    C->>C: Upload to Storage<br/>sage-audio-files/{userId}/{recordingId}.wav
    C->>C: Trigger Cloud Function

    Note over F: Cloud Processing
    F->>F: Parselmouth Analysis<br/>F0, Jitter, Shimmer, HNR
    F->>F: Write to Firestore
    F->>H: VocalBiomarkers<br/>(Real-time stream)
    H->>S: Comprehensive Results<br/>(30-60 seconds)
```

Domain-Driven Architecture

Core Domain Models

```
classDiagram
    class VocalBiomarkers {
        +F0Analysis f0
        +VoiceQualityAnalysis voiceQuality
        +VocalStabilityScore stability
        +ClinicalVoiceAssessment clinicalSummary
    }

    class F0Analysis {
        +Double mean
        +Double std
        +Double confidence
        +F0StabilityLevel stabilityLevel
        +isWithinClinicalRange(VoiceDemographic) Bool
    }

    class VoiceQualityAnalysis {
        +JitterMeasures jitter
        +ShimmerMeasures shimmer
        +HNRAnalysis hnr
        +VoiceQualityLevel overallQuality
    }

    class ClinicalVoiceAssessment {
        +VoiceQualityLevel overallQuality
        +F0StabilityLevel f0Stability
        +ClinicalRecommendation recommendedAction
        +String clinicalNotes
    }

    VocalBiomarkers --> F0Analysis
    VocalBiomarkers --> VoiceQualityAnalysis
    VocalBiomarkers --> ClinicalVoiceAssessment
    VoiceQualityAnalysis --> JitterMeasures
    VoiceQualityAnalysis --> ShimmerMeasures
    VoiceQualityAnalysis --> HNRAnalysis
```

Clinical Validation & Performance

F0 Accuracy Validation:

- **>95% correlation** with Praat reference implementation
- **Validation dataset**: 500+ clinical voice samples with known F0 values
- **Cross-platform testing**: iOS simulator vs device hardware differences
- **Real-time processing**: <5 seconds for local analysis, 30-60 seconds for comprehensive cloud analysis

Clinical Thresholds:

- **Jitter Local**: <1.04% Excellent, <2.5% Good, >2.5% Poor
- **Shimmer Local**: <3.81% Excellent, <6.5% Good, >6.5% Poor
- **HNR Mean**: >20dB Excellent, >15dB Good, <15dB Poor

- **F0 Stability**: >80% Excellent, >60% Good, <60% Poor

Quality Gate Architecture

Audio Quality Validation

```
flowchart TD
    subgraph "Audio Recording"
        A1[User Records Audio<br/>Sustained vowel or speech]
        A2[RMS Calculation<br/>Signal strength validation]
    end

    subgraph "Quality Gate Decision"
        B1{RMS >= Minimum?<br/>Device: 0.01<br/>Simulator: 0.005}
        B2{RMS >= Warning?<br/>Device: 0.006<br/>Simulator: 0.003}
    end

    subgraph "Analysis Outcomes"
        C1[❌ Reject Analysis<br/>insufficientSignalLevel error]
        C2[⚠️ Degraded Analysis<br/>30% confidence reduction]
        C3[✅ Normal Analysis<br/>Full confidence calculation]
    end

    A1 --> A2
    A2 --> B1
    B1 -->|No| C1
    B1 -->|Yes| B2
    B2 -->|No| C2
    B2 -->|Yes| C3

    classDef recording fill:#e3f2fd,stroke:#1976d2,stroke-width:2px
    classDef decision fill:#fff3e0,stroke:#f57c00,stroke-width:2px
    classDef outcome fill:#e8f5e8,stroke:#388e3c,stroke-width:2px

    class A1,A2 recording
    class B1,B2 decision
    class C1,C2,C3 outcome
```

Cloud Infrastructure

Firebase Architecture

```
graph TD
    subgraph "Firebase Ecosystem"
        A1[Cloud Storage<br/>Audio Files]
        A2[Cloud Functions<br/>Processing]
        A3[Firestore<br/>Results DB]
        A4[Authentication<br/>User Management]
    end

    subgraph "Audio Processing Pipeline"
        B1[main.py<br/>Entry Point]
        B2[feature_extraction_pipeline.py<br/>Orchestration]
        B3[vocal_analysis_extractor.py<br/>Parselmouth Integration]
    end

    subgraph "Clinical Models"
        D1[VoiceQualityLevel<br/>Assessment Categories]
        D2[F0StabilityLevel<br/>Stability Classification]
        D3[ClinicalRecommendation<br/>Action Items]
    end

    A1 --> B1
    B1 --> B2
    B2 --> B3
    B3 --> D1
    B3 --> D2
    D1 --> D3
    D2 --> D3

    classDef firebase fill:#fff3e0,stroke:#f57c00,stroke-width:2px
    classDef pipeline fill:#e3f2fd,stroke:#1976d2,stroke-width:1px
    classDef clinical fill:#e8f5e8,stroke:#388e3c,stroke-width:1px

    class A1,A2,A3,A4 firebase
    class B1,B2,B3 pipeline
    class D1,D2,D3 clinical
```

Data Flow & Real-time Updates

```
flowchart LR
    subgraph "iOS Client"
        A1[Voice Recording<br/>5s sustained vowel]
        A2[Local Analysis<br/>Immediate F0]
        A3[Upload Trigger<br/>Cloud processing]
    end

    subgraph "Cloud Storage"
        B1[sage-audio-files/<br/>{userId}/{recordingId}.wav]
        B2[Storage Trigger<br/>Function invocation]
    end

    subgraph "Analysis Processing"
        C1[Audio Validation<br/>Quality checks]
        C2[Parselmouth Analysis<br/>Research-grade extraction]
        C3[Clinical Assessment<br/>Threshold evaluation]
    end

    subgraph "Real-time Updates"
        E1[Firestore Listener<br/>VocalResultsListener]
        E2[UI Update<br/>Dashboard refresh]
        E3[Clinical Interpretation<br/>User-friendly display]
    end

    A1 --> A2
    A1 --> A3
    A3 --> B1
    B1 --> B2
    B2 --> C1
    C1 --> C2
    C2 --> C3
    C3 --> E1
    E1 --> E2
    E2 --> E3

    classDef client fill:#e3f2fd,stroke:#1976d2,stroke-width:2px
    classDef storage fill:#fff3e0,stroke:#f57c00,stroke-width:2px
    classDef processing fill:#e8f5e8,stroke:#388e3c,stroke-width:2px
    classDef realtime fill:#fce4ec,stroke:#c2185b,stroke-width:2px

    class A1,A2,A3 client
    class B1,B2 storage
    class C1,C2,C3 processing
    class E1,E2,E3 realtime
```

Testing Strategy

Comprehensive Test Coverage

```
graph TD
    subgraph "iOS Tests"
        A1[OnboardingTestHarness<br/>Voice recording flow]
        A2[AuthViewModelTests<br/>Authentication logic]
        A3[LocalVoiceAnalyzerQualityGateTests<br/>Quality gate validation]
    end

    subgraph "Quality Gate Tests"
        B1[Silent Audio Rejection<br/>RMS validation]
        B2[Good Quality Acceptance<br/>Above threshold validation]
        B3[Degraded Quality Handling<br/>Warning threshold handling]
    end

    subgraph "Cloud Function Tests"
        C1[test_feature_extraction_pipeline<br/>End-to-end processing]
        C2[test_vocal_analysis_extractor<br/>Parselmouth integration]
    end

    subgraph "Integration Tests"
        E1[End-to-End Flow<br/>iOS → Cloud → Result]
        E2[Real-time Updates<br/>Firestore listeners]
        E3[Error Recovery<br/>Network failures]
    end

    A1 --> E1
    A2 --> E3
    A3 --> B1
    A3 --> B2
    A3 --> B3
    B1 --> E1
    B2 --> E1
    B3 --> E1
    C1 --> E1
    C2 --> E1

    classDef ios fill:#e3f2fd,stroke:#1976d2,stroke-width:2px
    classDef qualitygate fill:#fff3e0,stroke:#f57c00,stroke-width:2px
    classDef cloud fill:#e8f5e8,stroke:#388e3c,stroke-width:2px
    classDef integration fill:#fce4ec,stroke:#c2185b,stroke-width:2px

    class A1,A2,A3 ios
    class B1,B2,B3 qualitygate
    class C1,C2 cloud
    class E1,E2,E3 integration
```

Performance & Scalability

Key Performance Metrics

```
graph LR
    subgraph "Performance Targets"
        A1[Local Analysis<br/>< 5 seconds<br/>SFVoiceAnalytics]
        A2[Cloud Upload<br/>< 10 seconds<br/>Firebase Storage]
        A3[Cloud Processing<br/>30-60 seconds<br/>Parselmouth]
        A4[Real-time Updates<br/>< 2 seconds<br/>Firestore Listener]
    end

    subgraph "Clinical Precision"
        C1[F0 Accuracy<br/>> 95% correlation<br/>with Praat reference]
        C2[Research Grade<br/>3 decimal places<br/>Clinical standards]
        C3[Audio Validation<br/>Duration + RMS + format<br/>Multi-layer quality checks]
        C4[Error Rate<br/>< 1% processing failures<br/>Robust error handling]
    end

    subgraph "Scalability"
        D1[Concurrent Users<br/>1000+ simultaneous<br/>Firebase autoscaling]
        D2[Daily Recordings<br/>10,000+ per day<br/>Cloud Functions]
        D3[Storage Growth<br/>1TB+ audio files<br/>Cost optimization]
        D4[Database Queries<br/>Sub-second response<br/>Firestore indexing]
    end

    A1 --> C1
    A2 --> C2
    A3 --> C3
    A4 --> C4

    classDef performance fill:#e8f5e8,stroke:#388e3c,stroke-width:2px
    classDef clinical fill:#e3f2fd,stroke:#1976d2,stroke-width:2px
    classDef scalability fill:#f3e5f5,stroke:#7b1fa2,stroke-width:2px

    class A1,A2,A3,A4 performance
    class C1,C2,C3,C4 clinical
    class D1,D2,D3,D4 scalability
```

Implementation Status

Core Components Status

Component	Status	Location	N
LocalVoiceAnalyzer	Complete	Domain/Services/LocalVoiceAnalyzer.swift	Workin analysi SFVoic + Qual
HybridVocalAnalysisService	Complete	Domain/Services/HybridVocalAnalysisService.swift	Full orc of loca analysi
Cloud Functions Pipeline	Complete	functions/ directory	Parsel analysi
VocalBiomarkers Models		Domain/Models/VocalBiomarkers.swift	Domai

	Complete		implem
Quality Gate Tests	Complete	SageTests/Domain/LocalVoiceAnalyzerQualityGateTests.swift	Compr test su
Domain-Driven Architecture	Complete	Organized in Domain/, Infrastructure/, Features/	Clean s of conc

Key Architectural Decisions

1. **Hybrid Analysis Approach**
    - Local iOS analysis for immediate feedback (< 5 seconds)
    - Cloud analysis for comprehensive research-grade features (30-60 seconds)
    - Progressive UI updates as results become available
2. **Dual Firestore Write Strategy**
    - Primary: recordings/{recordingId}/insights/ for canonical data
    - Secondary: users/{userId}/voice_analyses/{recordingId} for user-centric queries
    - Ensures backward compatibility and efficient querying
3. **Quality Gate Implementation**
    - RMS-based signal validation with platform-specific thresholds
    - Degraded analysis support for low-quality audio
    - Testing override capability for development scenarios
4. **Clean Domain Architecture**
    - Domain layer: Pure business logic and models
    - Infrastructure layer: External service integrations
    - Features layer: UI components with ViewModels
    - Shared layer: Cross-cutting concerns

Maintainers: This architecture document reflects the current implementation. The system follows clean domain-driven design principles with proper separation of concerns, improved testability, and better maintainability.