Noise Reduction Details

1. Noise Processing:

   Densper incorporates a speech recognition engine specifically optimized for the unique acoustic conditions found in dental clinics. To achieve stable and accurate performance in this high-noise environment, the system integrates three key elements:

   (1) a Speech Enhancement (SE) module,

   (2) models trained on noise-mixed clinical speech data, and

   (3) a Voice Activity Detection (VAD) module tuned to the characteristic noise patterns of dental procedures.

2. SE (Speech Enhancement): Technology for reconstructing clean speech from noisy input

   • Removes environmental noise, reverberation, and background interference to improve speech quality and intelligibility, enabling high-accuracy speech recognition results.

   • The Speech Enhancement model can be fine-tuned, allowing optimization for specific acoustic environments.

   • Dual-decoder architecture (magnitude and phase) enables high-fidelity speech reconstruction.

   • The service architecture achieves the same performance as transformer-based models while reducing computational load by approximately 70%

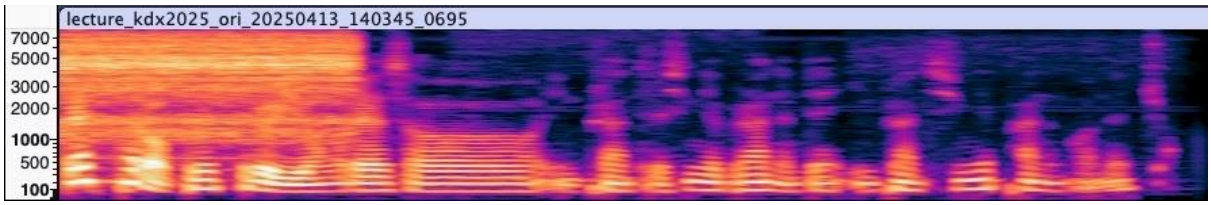

Figure 1. Spectrogram of actual speech containing noise

Noise reduction after applying the Densper SE module

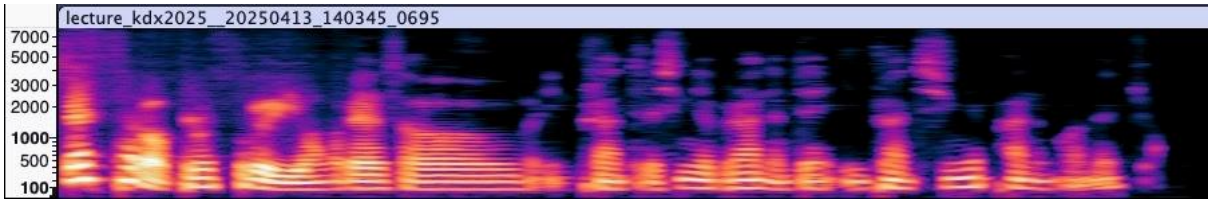

Figure 2. Spectrogram of speech after noise reduction

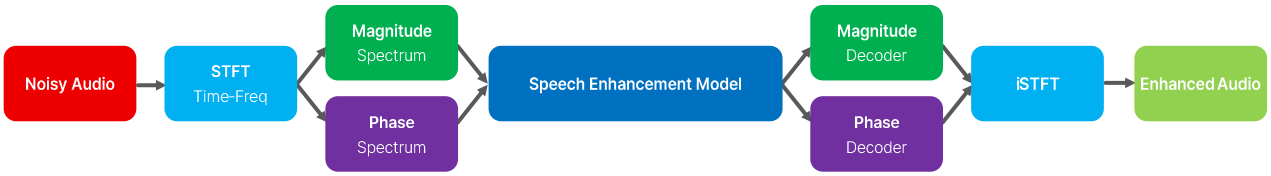

Figure 3. Speech Enhancement Pipeline

3. Noise-Robust Training: Incorporating dental machine noise and real clinical ambient noise into the model

• By employing a **Phase-Aware Decoding Architecture**, the model enhances noise robustness while preserving signal fidelity.

• **Complex Linear Layer** training enables simultaneous learning of both magnitude and phase components, preserving the full spectral representation compared to magnitude-only approaches.

• Strengthened **temporal coherence** allows accurate reconstruction of the attack/decay envelope and improves onset detection accuracy by approximately **5–10%**.

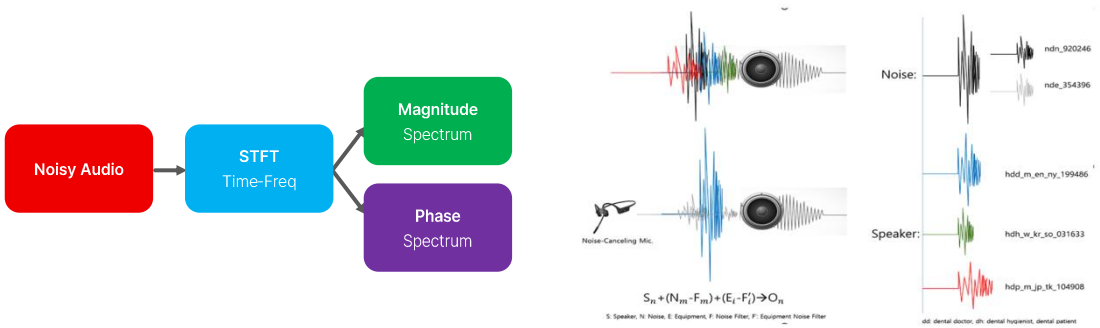

Figure 4. Concept of audio separation based on magnitude and phase

4. VAD Module: AI-based detection of speech segments

• Removes unnecessary silent segments, improving speech recognition accuracy while optimizing model resource efficiency and overall performance.

• The VAD module is trained on noise patterns commonly observed in clinical dental environments, ensuring robust speech detection and recognition performance during real procedures.

• Densper supports fine-tuning of the VAD module, enabling environment-specific optimization.

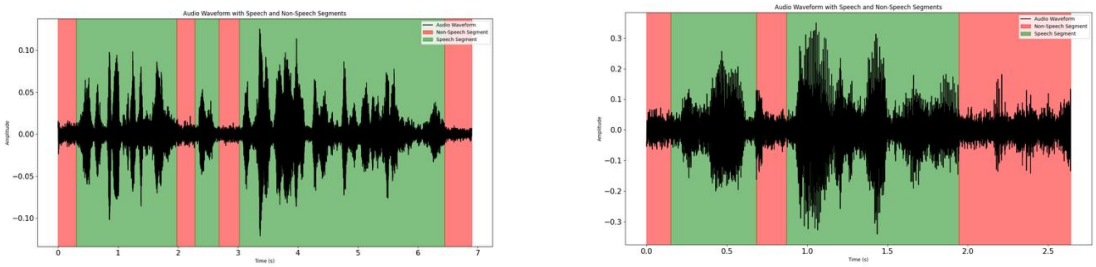

Figure 5. Silence detection comparison: Standard VAD vs. Densper VAD

As illustrated in *Figure 5. Silence Detection Comparison: Standard VAD vs. Densper VAD*, a conventional VAD tends to misinterpret dental clinic noise as valid speech, resulting in a high likelihood of recognition errors in real clinical settings. In contrast, Densper's fine-tuned VAD module accurately identifies and suppresses noise specific to dental environments, effectively treating it as silence and thereby preventing misrecognition.

5. Conclusion

By selecting the most effective model and configuration method across the three core modules described above, Densper's speech recognition pipeline achieves both high recognition accuracy and efficient engine resource utilization. One of the key advantages of AIzac's technology is that all essential components are developed in-house, which allows us to directly fine-tune and optimize each module based on the specific acoustic and operational conditions of the customer environment.