ADL

solution brief

Extract the maximum benefit from voice

Yobe's ADL software solution enables your organization to extract the maximum benefit from voice -- for speech recognition, voice biometrics, voice analytics, and more.

A proprietary, on-the-edge, artificial intelligence engine, Yobe's ADL (**Adaptive Discriminant Listening**) software solution effectively listens for voice (or other audio sources) in complicated audio settings, at a level of performance and scalability not previously seen. Through its improved signal and advanced insights, ADL enables your organization to extract the maximum business and customer benefit from voice -- for speech recognition, voice analytics, and more.

Key features and benefits include:

Edge Processing

Operates 100% on the edge with no need for Internet or cloud computing.

No Unwanted Artifacts

Processing does not introduce unnatural artifacts that adversely affect ASR and other voice analytics platforms

Track Voice of Interest

Separates and tracks direction-of-interest voice from other voices and sounds

No Reference Signal

Operates with no a-priori information about the auditory scene.

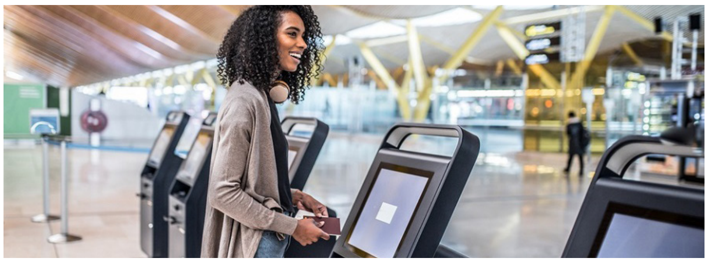
Co-Directional Effective

Our proprietary aging solution is able to separate co-directional sources in scenarios where beamforming is not effective.

Wind Noise Adaptive

Solution utilizes proprietary adaptive algorithms that allow accurate voice capture in outdoor and environmentally challenging audio scenarios.

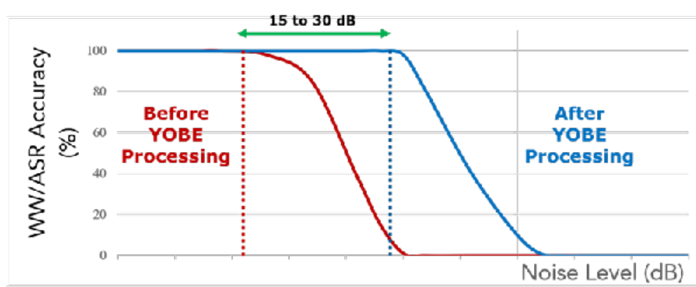
ADL gives device manufacturers, application developers, brands, and speech partners the ability to deliver better business outcomes and customer experiences across a **wide range of industries and use cases**, including mobile phones, smart speakers, appliances, in-car use, retail kiosks, security, and other applications



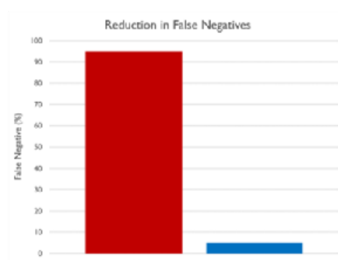
Focus on the voice, not the noise

ADL isolates the voice of interest and lowers the noise level, enabling your business to focus on the audio and outcomes that matter most.

ADL isolates the voice of interest and lowers the noise level by at least 15 dB and by up to 30 dB, with excellent ASR/KWR results and processing latency as low as 50ms. Compared to conventional solutions, ADL yields significantly lower error rates for machine-based recognition, enabling better results for your business. In the chart to the right ASR (Automatic Speech Recognition) and Keyword Recognition (KWR) accuracy were evaluated on real-world scenarios with and without ADL. The noise level at the device remained consistent at 70db and the voice of interest Signal-to-Noise Ratio (SNR) was adjusted from +5 dB to -25dB in 5dB increments.



False triggers reduced by 74%



Missed triggers reduced by 95%

Intelligent listening, on the edge

ADL intelligently listens, distinguishing between simultaneous audio sources, all in an on-the-edge solution.

The key to ADL performance is its ability to act as a **perceptual discriminant for distinguishing between simultaneous signal contributions** of multiple audio sources and producing an artifact-free, authentically reconstituted signal in which the desired voice signal is pushed to the auditory foreground while the remaining source contributions are relegated to the auditory background. The perceptual discriminant even enables separation of sources emanating from the same direction (and thus not separable via beamforming) by using **Yobe's aging analysis of sounds, a proprietary combination of AI, acoustics, and signal processing**. ADL adaptively utilizes evidence for source category (e.g., a voice), source location (e.g., within a local perimeter), source origin (e.g., a biometric identifier), source constitution (e.g., a pitch track), source kinetics (e.g., movement toward array), and source state (e.g., an indication of anger) to discern spatial data (direction, range etc.) and perceptual information (identity, emotional state etc.) for the voice of interest in complex and dynamically changing auditory scenes.

ADL enhances the desired signal to disambiguate spatial data (for items such as orientation and range) and to discern perceptual information (for items such as identity and emotional state). ADL is a library that runs on the edge - without requiring an internet connection or cloud-based computing, and can operate on a variety of hardware platforms and operating environments. At the core of the Yobe ADL solution is an **auditory intelligence capability** enabled through an inferential AI algorithm coupled with proprietary signal and perceptual information processing. Its real-time, on-the edge implementation uses unsupervised machine-learning principles to achieve discriminant listening objectives without the necessity of extensive training on huge audio databases or prior knowledge of the specific auditory scene.

The right answer for your use case

ADL has different flavors based on your voice use case. The key questions are: how many microphones and where are the noises?

ADL varies based on how many microphones signals are available for use. When two microphones are used, we can gather tremendous insights, and the solution is agnostic to the positions of the microphones. When more than two microphones are used, there are more audio channels from which to gather insights, and by factoring in the positions of the microphones - an even greater level of insight can be achieved.

ADL also varies based on where the noises are, i.e. what type of sound shielding is needed. If the dominant noises of concern are closer to the microphones than the voice of interest is, then LSS (Local Sound Shielding) is used - to shield out sounds originating from inside a spherical region around the microphones. However, if the dominant noises of concern are farther from the microphones than the voice of interest is, then PSS (Panoramic Sound Shielding) is used - to shield out sounds originating from outside a spherical region around the microphones. Lastly, if the dominant noises of concern may be either closer to or farther from the microphones than the voice of interest is, then DSS (Dynamic Sound Shielding) is used - to dynamically switch between LSS and PSS while tracking the voice of interest and noises, as the sound (and maybe microphone) positions change.

Minimum Requirement	
Data Memory	200 kbytes
Library Size	200 kbytes
1-CPU Usage	15%
Platform	Linux, Windows, ARM (32-bit, 64-bit)

Metric	Results
Processing Latency	less than or approx. 50 ms
SNR Improvement	Up to 30 db
ASR/STT accuracy of 95%+	Noise up to 4x louder than voice of interest
Keyword recognition accuracy of 95%+	Noise up to 8x louder than voice of interest