The Case for Direct- Sound- Enhancement using Acoustic- Proximity- Tuning

An SDA Consulting White Paper

A white paper on designing sound systems that improve speech intelligibility by maximizing direct sound at the listener

Executive Summary

Speech intelligibility in reinforced spaces depends less on how loud a system is than on what reaches the listener's ears first, and in what proportion to everything else. This paper sets out the case for D**irect-Sound-Enhancement through Acoustic- proximity- tuning**: a design approach that shortens the effective acoustic path between the sound source and each listener, controls where energy goes and aligns arrivals in time. Together these raise the direct-to-reverberant ratio (DRR), the most important acoustic factor in intelligibility. The paper explains the underlying acoustics, describes the design strategies, and gives a practical method for specifying, tuning and verifying such a system.

1. The Problem: Why Speech Fails in Real Rooms

In most rooms, a listener hears three kinds of sound:

- **Direct sound**, which travels straight from the loudspeaker to the listener.
- **Early reflections**, which arrive within about 50 ms and generally support intelligibility by fusing with the direct sound.
- **Late reflections and reverberation**, which arrive later, smear the rapid transitions between speech sounds, and mask the next syllable.

It is often said that consonants carry most of the information in speech. Work by Sungyub Yoo and colleagues at the University of Pittsburgh points to a more precise picture. They separated speech into quasi-steady-state parts, meaning the sustained portions of vowels and consonants, and transient parts, meaning the rapid transitions between sounds. The transient component held only a small fraction of the energy, yet it was almost as intelligible as the original speech (Yoo et al., 2004; 2007). The critical information sits in these brief, low-energy transitions, not in one class of phoneme.

That is exactly what late energy destroys. When late reflections are strong relative to direct sound, they smear the transitions and fill the gaps with sustained, high-energy vowel sound. Simply turning the system up does not help, because reverberant energy increases at the same rate as direct energy. Only changing the *proportion* helps.

2. The Core Principle: Proximity and the Direct-to-Reverberant Ratio

2.1 The inverse square law versus the reverberant field

Direct sound falls by about 6 dB each time the distance doubles. The reverberant field, by contrast, is roughly uniform across a room. The point where the two are equal is the **critical distance** (Dc). Inside Dc,

direct sound dominates and intelligibility is usually good. Beyond it, reverberation dominates and intelligibility falls quickly.

2.2 What determines critical distance

Critical distance grows with:

- The **directivity factor (Q)** of the loudspeaker. Doubling Q increases Dc by about 41%.
- The **total absorption** in the room. More absorption means less reverberant energy.

It shrinks as room volume increases with little absorption, which is typical of churches, atriums, transport halls and gymnasiums.

2.3 The proximity argument

Acoustic- proximity- tuning rests on one idea: **every listener should sit within the critical distance of at least one primary source**, and the energy that source sends to non-listening surfaces should be kept to a minimum. There are three ways to achieve this. Move the sources closer to listeners, raise their directivity so Dc reaches further, or reduce reverberant energy. A good design uses all three.

2.4 The listener's task: a timed multiple-choice test

Dominic Massaro's research on speech perception, including the Fuzzy Logical Model of Perception, treats recognition as a process of weighing evidence from several sources and choosing among candidate alternatives (Massaro, 1987; 1998). Building on that framework, this paper offers an analogy: listening to, understanding and remembering speech is something like passing a timed multiple-choice test. For each syllable and word, the listener has to choose among the possible answers using whatever auditory, visual and contextual evidence is available, and do it before the next item comes along.

This view changes how we judge a sound system. The system is there to help the listener pass that test:

- **Clear evidence per item.** Strong direct sound gives each speech sound sharp cues, so fewer alternatives remain plausible.
- **No interference between items.** Late reverberation spills one syllable into the next, which is like the previous question bleeding onto the current page.
- **Time to answer.** Each extra effort spent decoding takes working memory away from understanding and remembering. Clean speech leaves more capacity for meaning and recall.
- **More than one source of evidence.** Massaro's research shows that listeners combine what they hear with what they see. Seeing the talker's face adds independent evidence, especially for syllables that sound alike but look different.

Proximity tuning improves the auditory evidence. The audio-visual anchor described in section 3.8 makes sure that evidence lines up with the visual evidence.

2.5 Griesinger: proximity, phase coherence and the audibility of direct sound

Acoustician David Griesinger argues that clarity depends on whether the ear can **detect the direct sound** in the brief onset of each syllable, before reflections bury it. He ties the sense of auditory **proximity**, meaning that a talker sounds close and engaging, directly to intelligibility and recall (Griesinger, "The physics of auditory proximity and its effects on intelligibility and recall").

The key findings for this paper:

- **Phase coherence of harmonics.** Our ability to separate a talker from noise and from other voices, and to judge direction and distance, depends on the phase relationships between the harmonics of the voice, especially in the vocal formant range above about 1 kHz. Reflections scramble these phases in predictable ways (Griesinger, "The audibility of direct sound as a key to measuring the clarity of speech and music").
- **Early reflections can hurt.** The usual view is that early reflections help intelligibility. Griesinger finds that they can randomize harmonic phases too and so remove the ability to separate sound streams.
- **A threshold, not a gradient.** In a given room there is a fairly sharp distance beyond which localization and proximity disappear. He calls this the Limit of Localization Distance (LLD), and beyond it speech becomes hard to localize, harder to understand and harder to remember (University of Rochester lecture).
- **The LOC measure.** Griesinger developed LOC, a measure taken from a binaural impulse response that compares the strength and timing of the direct sound with the reflections that follow in roughly the first 100 ms. It predicts whether the direct sound will be audible at that seat (ASA lay-language paper).
- **Direct sound needs to compete with total reflected energy.** In small rooms, the direct sound must be at least as strong as all the reflected energy for accurate imaging, and the threshold falls as the room gets larger (AES 126th Convention paper).

What this means for design: proximity tuning should put every listener inside the LLD of a source, not just inside the classic critical distance. It should also limit strong early reflections from nearby surfaces, not only the late reverberation. It links to Massaro's timed test as well: a coherent direct sound makes each "answer" quick and certain, which leaves working memory free for recall. LOC is also worth measuring along with STI when the system is verified.

2.6 Engagement and selective attention: locking on to the direct sound

Intelligibility is not just about hearing words. The listener has to **engage**: their auditory system must pick out one source, the talker, and keep attending to it while ignoring noise, other voices and the room. This is the ability behind the cocktail party effect, and it depends on being able to lock on to the direct sound.

How it works, following Griesinger's account ("The Relationship Between Audience Engagement and Our Ability to Perceive the..."):

- **Direct sound is the handle for attention.** A coherent direct sound gives the talker a clear position, timbre and distance. The brain uses these cues to group that voice into one stream and hold attention on it.
- **Near sounds demand attention, and far sounds can be ignored.** People sort sounds almost at once into near and far. A voice heard as close draws attention and feels engaging. A voice heard as distant and smeared is treated as background, even when it is loud enough.
- **Reflections break the lock.** When reflections scramble the phase of the voice's harmonics, the cues that hold the stream together weaken. Attention drifts, effort goes up, and recall falls.

- **Engagement costs less when the lock is strong.** With a firm lock, selective attention runs almost automatically. Without it, the listener has to work to stay with the talker, which uses up the same limited capacity they need for Massaro's timed test.

What this means for design:

- Give every listener a direct sound strong and coherent enough to lock on to, from close range and with tight pattern control (sections 3.1 and 3.2).
- Make the locked source match the talker in place and time, using the audio-visual anchor (section 3.8), so that the ears and eyes point attention at the same place.
- Avoid several loudspeakers arriving at similar levels and times, because competing images make the brain choose between streams.
- Keep background noise low, because noise competes for the same attention.

3. Design Strategies for Direct Sound Enhancement

3.1 Distributed and near-field loudspeaker layouts

Instead of one central cluster trying to cover a deep room, a distributed design uses multiple smaller sources, each covering a limited zone at short range. Examples include:

- Ceiling loudspeakers at close spacing in low-ceiling rooms.
- Pew-back or seat-back loudspeakers in highly reverberant worship spaces.
- Under-balcony and front-fill loudspeakers to cover areas a main system cannot reach directly.
- Column array loudspeakers mounted on pillars along the length of a nave or hall.

Each listener is now perhaps 2 to 5 meters from a source rather than 20 to 30, which can raise the DRR by 10 dB or more.

3.2 High-directivity sources

Pattern control keeps energy on the audience and off walls, ceilings and glass. Useful tools include:

- **Pre-configured line arrays** and **digitally steerable columns**, whose vertical beam can be selected or aimed electronically to match the seating plane.
- **Constant-directivity horns**, which hold a consistent pattern across the speech band.
- **Beam shaping** to produce an even level from front to back while avoiding reflective surfaces.

Steerable columns are particularly valuable in heritage buildings, where they deliver high Q from a slim, visually discreet unit that can be made to work when they must be mounted at a non-optimum height.

3.3 Time alignment and the precedence effect

When several sources serve one listener, their arrivals have to be coordinated. Under the **precedence (Haas) effect**, the ear localizes to the first arrival and fuses later arrivals that come within about 5 to 30 ms, provided they are not much louder. Design rules:

- Delay each distributed loudspeaker so its sound arrives **5 to 10 ms after** the sound from the main source or the live talker. This keeps localization natural while the near source supplies the clarity.

- Avoid arrivals more than about 50 ms apart at similar levels, because they will be heard as echoes.
- Within each zone, use overlap and level shading so that no listener gets competing arrivals of equal strength.

3.4 Coverage uniformity

Intelligibility should be consistent, not just good on average. Aim for a direct-field level variation of no more than about ±3 dB across the listening area in the 2 kHz and 4 kHz octave bands, where much of the transient speech information is carried.

3.5 Spectral tuning for speech

- Apply high-pass filtering (typically 80 to 120 Hz for speech-only systems) so low-frequency energy does not excite room modes and add masking.
- Shape the response with a gentle emphasis on 1 to 4 kHz to support the transient cues in speech.
- Correct for the off-axis behavior of the loudspeakers, not just their on-axis response.

3.6 Supporting room acoustics

Electro-acoustic design works best alongside acoustic treatment. Placing absorption on the surfaces most likely to return late, strong reflections, such as rear walls and high ceilings, raises Dc for every source in the room.

3.7 Controlling background noise

A high signal-to-noise ratio is the second pillar of intelligibility. Aim for speech at least 10 to 15 dB above ambient noise. Proximity helps here as well, because a nearby source reaches the needed level with less total output and therefore excites less reverberation.

3.8 The audio-visual anchor loudspeaker

An A**udio-visual Anchor** is a loudspeaker placed close to the talker, at the lectern, pulpit or presenter position, so that the first sound to reach listeners comes from where they see the talker. Its job is to tie the auditory image to the visual image, so the listener can combine hearing and seeing without effort.

Why it matters:

- **Localization follows the first arrival.** Because of the precedence effect, if the anchor's sound arrives first, listeners hear the voice coming from the talker even when most of the energy comes from distributed loudspeakers nearby.
- **Audio-visual integration needs agreement.** When sound and lip movement come from the same place and at the same time, the brain fuses them into one stronger percept. When they come from different directions, attention is split and the visual benefit shrinks.
- **Synchrony.** Sound should not lag far behind the visible lips. Keep the total delay at any seat within about 40 ms of the visual event and avoid sound arriving before the light.

Design guidance:

1. Mount the anchor at or just above the talker's head height, close to the talker, using a compact unit with controlled directivity so it does not cause feedback.
2. Set its level so it provides localization for the front and middle seats without dominating the room.
3. Align every distributed and delayed loudspeaker to the anchor, so their sound arrives 5 to 10 ms after it at each seat.
4. Give the talker good lighting and a clear sightline so listeners can actually see the face; the anchor only helps if the visual cue is there.
5. Where a talker moves around, use several anchor positions or a tracking delay system so the sound image follows them.

3.9 Upward masking: a hidden risk with distributed ceiling loudspeakers

Upward spread of masking means that lower-frequency sound masks higher frequencies much more than the other way round, and the effect grows quickly as level rises. In speech, strong low and low-mid vowel energy can hide the weaker, brief 2 to 4 kHz transient cues that carry intelligibility (section 1), and that also carry the harmonic detail the ear uses to lock on to the direct sound (sections 2.5 and 2.6).

Why distributed ceiling systems are prone to it

- **Low frequencies add up and high frequencies don't.** Adjacent ceiling loudspeakers overlap widely at low and mid frequencies, so their output sums into a build-up. At high frequencies they are more directional, so a listener between or off-axis from loudspeakers gets mostly the lower band.
- **Boundary loading.** Flush-mounted ceiling loudspeakers gain low-frequency output from the ceiling surface, which tips the balance further.
- **Stacking on the main system.** A downstream or delayed loudspeaker adds its low-frequency energy on top of the low end already arriving from the anchor and main system, plus the room's reverberant low end.
- **Turning it up makes it worse.** Raising the level to fix poor clarity increases upward masking, so the system gets louder and muddier at the same time.

The result is a system that seems full and loud but sounds distant and hard to follow, which is the opposite of proximity.

Frequency shaping and shading for downstream loudspeakers

A downstream loudspeaker should deliver the **proximity cues, not the bulk of the energy**. The anchor and main system provide the low end and localization, and the downstream loudspeaker adds the clear, coherent upper band that lets the listener lock on.

- **High-pass higher than the main system.** Roll off downstream and ceiling loudspeakers well above the main system's high-pass. A starting point of about 300 to 600 Hz is common, adjusted by listening and measuring, so they add little to the low-mid build-up.
- **Shape for the transient band.** Keep the response focused on roughly 1 to 4 kHz, where the transient speech cues and the phase-coherent harmonics Griesinger describes live. Cut the low-mid region where overlap and boundary gain pile up.

- **Shade the level by zone.** Set each zone only as loud as it needs to be to bring that area's direct sound above its threshold, and taper levels across zone boundaries so no seat gets two strong, equal arrivals.
- **Combine shaping with time alignment.** Delay each zone to the anchor (section 3.3), so the added upper-band clarity merges with the talker's image instead of pulling attention away.
- **Choose loudspeakers that keep a consistent pattern.** Loudspeakers with controlled, consistent directivity across the speech band reduce the off-axis low-frequency imbalance in the first place.
- **Check the result at the seat.** Measure the summed spectrum at listening positions, not the response of a single loudspeaker on its own axis and confirm that STI and LOC improve rather than just level.

Applying the Approach in Churches and Lecture Halls

Churches and worship spaces

Traditional churches often have reverberation times of 2 to 4 seconds or more, hard stone surfaces and long naves, so a central system rarely keeps the back pews inside the critical distance.

- Use **column- arrays** on pillars along the nave, or **pew-back loudspeakers** in very reverberant buildings, so each zone is served from close range.
- Place an **audio-visual anchor** at the pulpit and another at the lectern or altar, and or crossfade between them as the speaking position changes.
- Delay each column zone back to the active anchor position.
- Keep loudspeakers discreet and finished to suit the architecture, and respect heritage constraints.
- Leave the room's natural reverberation for music and use the electro-acoustic system to keep speech clear.

Lecture halls

Lecture halls are usually less reverberant, but listeners must understand and remember dense information for an hour or more, which makes Massaro's timed-test view especially relevant.

- Use a **main system or anchor near the lectern**, with **delayed ceiling or under-balcony loudspeakers** for the rear rows in deep, raked halls.
- Give the lecturer's face good lighting and make sure any projected camera image of the talker is in sync with the audio.
- Keep background noise from ventilation and projectors low to protect the signal-to-noise ratio.
- Support assistive listening systems, such as hearing loops or infrared, fed from the same speech-optimized signal.
- Aim for STI of 0.60 or higher, given the heavy cognitive load of learning.

4. A Practical Design and Tuning Method

1. **Assess the room.** Measure or model reverberation time (RT60) by octave band, background noise and room volume. Work out the critical distance for candidate loudspeaker types.

2. **Define listening zones.** Map the seating and standing areas and identify where listeners would sit beyond Dc of any central source.
3. **Select and place sources.** Choose loudspeaker types and positions so every zone falls inside Dc, with directivity aimed at listeners and away from reflective surfaces.
4. **Model predictively.** Use acoustic simulation software to predict direct-field coverage, DRR and STI before installation.
5. **Set delays and levels.** Align each delayed zone to the main source or talker position using measurement, and add the precedence offset.
6. **Equalize and filter.** Tune for a smooth, speech-focused response at representative seats.
7. **Verify.** Measure the Speech Transmission Index (STI or STIPA) and, from binaural impulse responses, Griesinger's LOC on a grid of seats and confirm coverage uniformity and echo-free arrivals.
8. **Document and train.** Record settings and give operators clear guidance on gain structure and microphone technique.

5. Performance Targets

Commonly cited targets for speech reinforcement are:

Metric	Target
STI (general speech)	0.50 or higher; 0.60 or higher preferred
STI (life-safety voice alarm, typical standard)	0.50 or higher
Clarity C50	0 dB or higher
Signal-to-noise ratio	10 to 15 dB or more
Direct-field uniformity, 2 to 4 kHz	about ±3 dB

Always check the local standards and codes that apply to the project, as specific requirements vary by jurisdiction and application.

6. Benefits of the Approach

- **Higher intelligibility** in rooms where central systems cannot succeed.
- **Lower overall sound pressure**, which means less listener fatigue and less spill into nearby spaces.
- **Better accessibility** for people who are hard of hearing, older listeners and non-native speakers, who are all more sensitive to reverberation.
- **Natural localization**, because correctly delayed systems keep attention on the talker.
- **Architectural sympathy**, since smaller, discreet sources suit heritage and design-led interiors.

7. Limitations and Trade-offs

- More loudspeakers mean more cabling, amplifier channels and installation cost.
- Delay alignment must be accurate for a defined zone or listeners near zone boundaries may hear some imprecision.

- Very close sources can make the sound seem to come from the loudspeaker rather than the talker if levels are not carefully controlled.
- Ongoing maintenance and operator discipline are needed to keep performance from drifting.

8. Conclusion

Intelligibility is decided by the balance of energy at the listener's ear, not by system power. By bringing sources acoustically closer to listeners, controlling directivity, aligning arrivals in time and supporting the result with room treatment and noise control, designers can put every listener inside the critical distance and deliver clear, natural speech even in difficult spaces. Direct sound enhancement through acoustic proximity tuning is therefore the most reliable foundation for speech reinforcement in reverberant environments.

Glossary

- **Critical distance (Dc):** The distance from a source where direct and reverberant levels are equal.
- **Direct-to-reverberant ratio (DRR):** The level of direct sound relative to reverberant sound at a given position.
- **Directivity factor (Q):** How strongly a loudspeaker concentrates energy in one direction.
- **Precedence (Haas) effect:** The ear's tendency to localize to the first-arriving sound.
- **STI:** Speech Transmission Index, an objective intelligibility measure from 0 (bad) to 1 (excellent).
- **C50:** The ratio of energy arriving in the first 50 ms to energy arriving later, used as a speech clarity measure.

References

Griesinger, D. (2009). *The importance of the direct to reverberant ratio in the perception of distance, localization, clarity, and envelopment.* Audio Engineering Society 126th Convention, Munich. PDF

Griesinger, D. *The audibility of direct sound as a key to measuring the clarity of speech and music.* Paper 1aAA14, Acoustical Society of America meeting. PDF; lay-language summary

Griesinger, D. *The relationship between audience engagement and our ability to perceive the pitch, timbre, azimuth and envelopment of multiple sources.* Presentation. PDF

Griesinger, D. *Lecture on auditory proximity, intelligibility and recall, including the Limit of Localization Distance (LLD).* University of Rochester. PDF

Griesinger, D. (2022). *The physics of auditory proximity and its effects on intelligibility and recall.* VSA Oberlin First Fridays Online Seminar, 4 March 2022. Author's site

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle.* Cambridge, MA: MIT Press.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review, 85*(3), 172–191.

Yoo, S., Boston, J. R., Durrant, J. D., Kovacyk, K., Karn, S., Shaiman, S., El-Jaroudi, A., & Li, C.-C. (2004). Relative energy and intelligibility of transient speech components. *Proceedings of EUSIPCO 2004*, Vienna, 1031–1034. PDF

Yoo, S., Boston, J. R., El-Jaroudi, A., Li, C.-C., Durrant, J. D., Kovacyk, K., & Shaiman, S. (2007). Speech signal modification to increase intelligibility in noisy environments. *Journal of the Acoustical Society of America, 122*(2), 1138–1149. Abstract

Note: the timed multiple-choice analogy in section 2.4 is SDA Consulting's interpretation of Massaro's framework, not a direct quotation.

© 2026 SDA Consulting. All rights reserved.