

The "BRAIN" Model of Intelligibility in Business Telephony

Jeff Rodman

Fellow/CTO

September 2006

"I am in susquehanna jail. Susquehanna.
S – u – s – q –
'Q!' 'Q,' you know, the thing you play billiards with.
Billiards. B – i – l – l – i –
No, 'L', for 'Larynx!'
L – a – r – y – n –
No! Not 'M!' 'N'!!!
'N', as in 'Neighbor!' N – e – i ..."

---Eric Blore, in jail and on the phone to Fred Astaire in "Shall We Dance," 1937

Introduction

There are times when there is no substitute for being understood on the telephone. Well before 1937, the limitations of the telephone in accurately conveying speech were known. Irregular and limited bandwidth, noise, variations in end-to-end loudness, sidetone and distortion had all been identified as contributors to degradation of the spoken word. In 1910, Campbell performed experiments in which he found 59 percent accuracy when words were called over the telephone, as compared to 96 percent through open air. The abilities of the telephone as an efficient and accurate channel for human speech have always been regarded with a bit of a wink and a chuckle, tolerated due to a common understanding that it is the best available.

Despite modern digital trunking and switching technology, this remains an everyday problem. Why? Analog loop lengths, building wiring, variable line equalization characteristics, poor handset and speakerphone designs, mixed networks, noise in conference rooms, paper shuffling, pen tapping, fan noise, and a host of other issues are still with us, even here in the digital age.

This paper discusses these issues and presents the "BRAIN" model of critical elements in business telephony. We show how their mutual dependencies can be used to improve telephone and audio system performance, and how contemporary systems can produce direct benefits to clear communications.

What Distinguishes Business Telephony?

In business, the critical role of the telephone is magnified for a number of reasons. Consider the following characteristics of business telephony.

- Time is often in more demand than in personal telephony, which makes misunderstandings and "can you repeat that?" more frustrating. Meetings have

fixed lengths, so lost time is irretrievable.

- The nature of discussions makes accuracy critical, so the many ways that telephones distort speech and degrade accuracy carry a real cost to business.
- Users are often talking to people that they do not know and have never met before. A phone call is the first impression for both parties in a high-value relationship.
- Phone calls often occur between people who have different native languages or dialects, so accented speech is added to the other burdens of telephony.
- Conferences occur among groups, which increases both the cost and the potential value of the meeting due to the number of participants and the difficulty of scheduling them. Yet, to accommodate such a group, the quality of the sound must be degraded because some kind of speakerphone is required. This can introduce room reverberation, fan noise, clipping and feedback, interruptions, and multiple participants sitting near and far away.
- Having multiple talkers participating in a conference makes fast and accurate identification of who is talking more important, but also more difficult.
- Meetings can be very long, yet require sustained attention. This puts an increased strain on sound quality, because small differences in speakerphone performance add up to big differences in fatigue and attention.

Intelligibility and the BRAIN Model

For all these reasons, intelligibility (how easily speech is understood) is especially critical in business telephony. Yet perversely, the challenges to intelligibility are magnified in the group settings that are so common in business. Let us look at the components of speech intelligibility.

We hear and understand speech in three main stages, which are called here the physical, cognitive, and analytical. In the physical stage, speech is carried from the talker's mouth to the listener's ears, and the fidelity with which this is done is paramount. In the cognitive stage, the listener resolves ambiguities in what she's heard by applying simple reasoning, such as grammatical and accent rules, to the local context of the word. And finally, in the analytical stage, those words that are not resolved through the first two stages are subjected to more intense scrutiny, examining the troubling words in broader contexts to see if their identities can be inferred.

These second two layers of comprehension are increasingly distracting. Additionally,

there are instances such as speech from foreign talkers, where the assumptions of these contextual analyses do not hold: a person who does not share the same native language or dialect will likely use entirely different words or sounds than those that are expected.

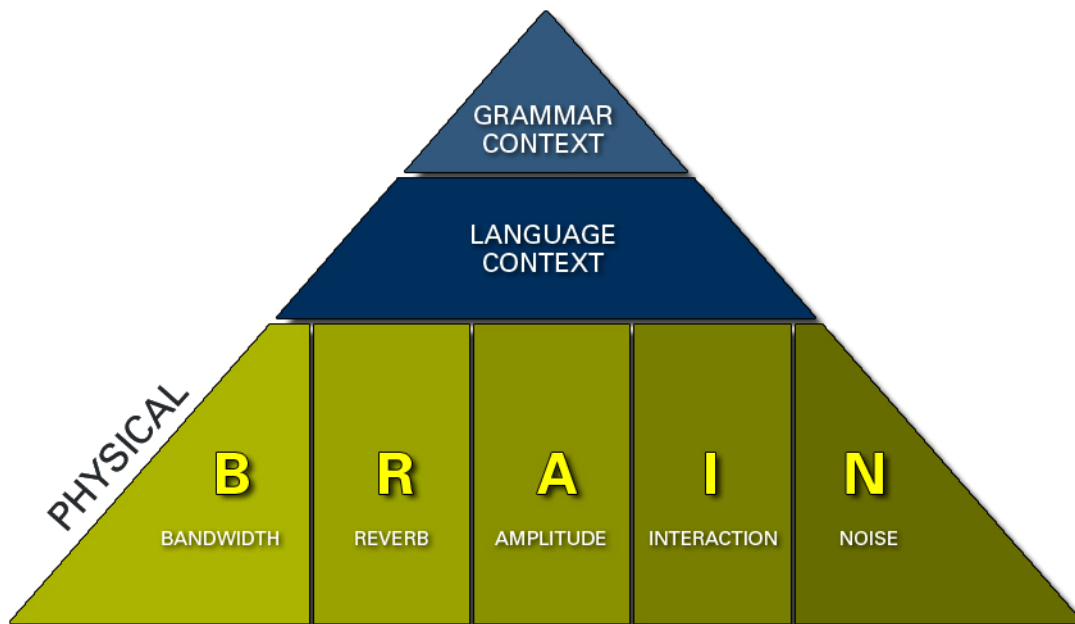


Figure 1. The BRAIN Model of Physical Speech Communication

For these reasons, the physical elements of speech are the most critical line of defense. If speech is clearly conveyed to the listener in the first place, remaining ambiguities will be less frequent and the meeting will proceed more smoothly.

The Physical Elements of Intelligibility

There are five critical parameters in the physical communication of speech: bandwidth, reverberation, amplitude, interaction, and noise. Together, these form the BRAIN model. These five parameters constitute the foundation upon which speech comprehension and talker identification rest in business audio communication.

Bandwidth is the amount of speech bandwidth that is carried to the listener. Telephones, which are limited to the band 300 Hz to 3.3 kHz, carry only 20 percent of the frequencies that are present in human speech. In contrast, some modern business audio systems, such as videoconferencing systems and next-generation telephones and speakerphones,, carry frequencies as high as 22 kHz.

Reverberation measures the amount of room echo that occurs between the talker and the microphone. Reverberation, which makes speech more difficult to understand, is

strongly affected by room characteristics (hard, reflective walls, floor, and ceiling), room size, and the orientation between the speaker and the microphone. If the microphone is not pointed at the speaker or is more distant, a greater proportion of the sound picked up by the microphone will be reverberation instead of direct speech, and the end result will be a decrease in intelligibility.

Amplitude refers to how loud the talker sounds to the listener. A quiet talker is more difficult to understand than a loud one, all things being equal, and a listener who is distant from the loudspeaker will have a harder time than when he is closer to it. The telephone network can compound this situation because telephone lines can have very different gains, varying in loudness by up to 20 dB (100 times) from connection to connection.

Interaction is the ability of two or more participants to naturally interact with each other in a telephone conference. It is essential that one talker be able to interrupt another without disturbing the flow of conversation, or the dialogue is stilted and unnatural. The thread of meaning is lost as the participants are diverted into the kind of "what?" "you go ahead," "no, you go ahead" exchange so often heard with mediocre speakerphones.

Noise refers to the proportion of ambient noise that is picked up along with speech. Room noise, such as air conditioning and projector fan noise, can be easily heard by microphones, and it plays a significant role in decreasing the intelligibility of speech. This has become a more significant issue in recent years, as LCD projectors (which often have small, noisy fans) have appeared on conference tables adjacent to conference room audio system microphones.

What is interesting about these five BRAIN elements is that they work together in creating intelligible speech and identifiable speakers. Research has shown that, to a surprising degree, each of these five parameters can compensate for deficiencies in another. An excess of noise, for example, which is accidentally picked up and added to a talker, will make the speech more difficult to understand, but can be compensated by increasing the bandwidth of the whole signal. High room reverberation, which makes the signal muddy and difficult to understand, can be compensated by increasing the bandwidth of the signal, the loudness, or both. Deficient loudness, perhaps caused by a talker who is too far from the microphone, can be compensated by reducing the amount of noise, increasing bandwidth, or reducing reverberation.

Methods for BRAIN Optimization

This interaction is fortunate, because it means that many problems that cannot be directly solved (making a room less reverberant, for example, may involve significant expense) can be assisted by improving parameters which are more accessible. They may not affect the issue directly (increasing bandwidth will not decrease room reverberation, in this example), but they can help achieve the desired result (it will make the talker more easily understood).

Let us examine today's toolbox to understand what options we have in making meetings more efficient and enjoyable. As before, we will examine these in the context of the BRAIN model to understand what can go wrong, and what can be done about it.

Bandwidth. A deficiency in bandwidth, most often caused by using narrowband Plain Old Telephone System (POTS) or IP connections to carry a telephone connection, has long been regarded as an insoluble problem because it is seemingly constrained by one hundred years of infrastructure. New technology, however, is making this problem disappear. Standards-compliant IP telephones and speakerphones are becoming available that take advantage of the available data bandwidth to deliver much higher audio bandwidth and fidelity. Products with Polycom HD Voice technology feature a frequency range of up to 7 kHz or greater, which is over twice the range of conventional phone lines. And even analog telephony, via Polycom's VTX 1000 technology, is doubling available bandwidth from the conventional 300 Hz to 3.3 kHz of telephony to a 80 Hz to 7 kHz frequency range that rivals the best video systems, and does this over conventional telephone lines.

Reverberation. Reverberation can be addressed directly, or through other elements of the BRAIN model. One of the best head-on approaches to the problem of reverberation is to optimize the placement of the microphone with respect to the talker. This can be done by using a multiple-microphone system which intelligently selects the microphone having the best pickup for each talker, such as the Vortex Installed Voice system, or the SoundStation VTX 1000 tabletop systems, by ensuring that the principal talkers are positioned closest to the microphone, or by using a personal microphone (such as the Polycom wireless microphone) for the principal talker. Room treatments, of course, are also options for a direct solution in these cases; mounting acoustic diffuser and absorbing elements on the walls, ceiling, corners, and so forth, can make a significant improvement in the performance of the room.

In situations where these options have been exploited to the extent possible, the BRAIN model shows that excessive reverberation can also be compensated by the other available elements. Increasing bandwidth, for one, significantly increases the intelligibility of reverberant speech; in one test, word accuracy increased from 52 percent to 80 percent when bandwidth was boosted from 4 kHz to 8 kHz in reverberant space.ⁱ Increasing amplitude is somewhat effective, as is reducing noise if the environment has excessive noise.

Amplitude. Amplitude, or loudness, when insufficient, can make it difficult to hear a talker. A 20dB drop in channel gain slashes one standard measure of intelligibility, the articulation index, from 80 percent to 30 percent.ⁱⁱ Low amplitude can be caused because the talker is too far from the microphone, or the listener is too far from the speaker, or because the path between is too lossy. Moving the talker and listener are obvious solutions, but not always practical. Systems that can automatically adjust microphone gain can greatly help in these situations; microphone AGC (automatic gain

control), when carefully designed in a telephony audio system, can be the source of large improvements.

Again, the other elements of BRAIN can be applied to help overcome a problem with low amplitude. Increasing bandwidth (such as by replacing conventional speakerphones with a pair of IP or VTX 1000 units with Polycom HD Voice technology), reducing reverberation (such as with the microphone selection techniques described above), and reducing ambient noise can all substantially compensate for low amplitude.

Interaction. Interactive speech between distant groups can be difficult to conduct for a number of reasons, including the most straightforward, the absence of a true full-duplex system that allows transparent interactive speech. The approaches suggested above are also worth considering: better microphone placement, wider bandwidth, stronger signal.

Sometimes, it will be found that poor interaction is caused by an excessive end-to-end delay of the audio signal. As one source of delay, the communications channel itself may be at fault. Although less common today, satellite connections have substantially longer connection delay than earth-based ones. This is not a problem in one-way communications such as broadcasts, but can bring catastrophe to an attempted live interactive conference. Another source of delay can be found in IP telephony systems which sometimes can have delays as long as 150 ms. Delay also can come from some conference bridges, either POTS or IP, which can insert substantial delay in their processing. Room reverberation can increase the apparent delay between talker and listener. In all these cases, careful design and evaluation is the best first step.

Noise. Common noise sources share much of the same spectrum with speech, and can be quite effective at masking it and making it difficult to understand. First, try to fix noise at the source. Move the microphones farther from air conditioner ducts, overhead projectors, coffee makers, and so on. If the microphones are directional, you can achieve some benefit by pointing them away from the noise sources (although this must be done in a manner that still leaves them pointing at the talkers).

If these direct approaches leave the situation wanting, increased bandwidth, again, can significantly improve intelligibility. By providing more of the talker's voice, the listener finds it easier to separate speech from noise and follow the conversation. Another characteristic of the enhanced-bandwidth Polycom HD Voice technology found in certain desktop IP and conference phones, incidentally, is that they eliminate noise from the telephone line, so traditional clicking, buzzing, hissing, and other analog telephony artifacts are removed.

In another supportive approach to this problem, a few systems today incorporate active noise reduction algorithms that analyze the microphone audio in time and frequency domains, and apply a sophisticated form of filtering that can reduce fan noise by 6 - 9 dB, while having little or no effect on voice. The Polycom Vortex and SoundStation

VTX 1000 are examples of such systems.

A combination of the approaches described here is often successful in resolving common noise problems.

Conclusion

Recent advances in each of the major areas of the BRAIN model, combined with the fact that each of these can support and compensate for the others, are making good audio conference installations much easier to achieve, and with much better result. Regrettably for Eric Blore and Fred Astaire, it will probably be awhile before such a facility is available in Susquehanna jail, but dramatic improvements in business audio quality are today available to the rest of us at reasonable cost.

ⁱ George A. Campbell, "Telephonic Intelligibility," Philosophical Magazine, 19, ser. 6 (1910): 158.

ⁱ P. W. Barnett, Overview of Speech Intelligibility, Proceedings of the Institute of Acoustics, Rayleigh House, Bush Hill Park, Vol. 21 Part 5 (1999).

ⁱⁱ Harvey Fletcher and Rogers H. Galt, "The Perception of Speech and Its Relation to Telephony," The Journal of the Acoustical Society of America, 22 No. 2 (March 1950): 138.