Proc. of the 2017 International Symposium on Electromagnetic Compatibility - EMC EUROPE 2017, Angers, France, September 4-8, 2017

Multiple Sources in a Reverberant Environment: The "Cocktail Party Effect"

Perry F. Wilson
RF Technology Division
National Institute of Standards and Technology
Boulder, CO, USA
pfw@nist.gov

***Abstract* — Acoustics and electromagnetics have many parallels based on similar equations describing propagation, reflection, and reverberation. This paper considers one example, the "cocktail party effect", that describes speech comprehension in the presence of multiple background conversations. An electromagnetic equivalent is developed that may have application to describing the quality of a wireless communications link in the presence of multiple and similar background signals.**

Keywords—acoustics; coexistence; reverberation chamber; wireless communications

I. Introduction

The "cocktail party effect", coined by an acoustics researcher in 1953 [1], is a well-known phenomenon describing what happens when many people have multiple conversations in a reverberant room [2-5]. As the number of conversations increases and the ambient noise in the room rises, people in individual conversations will tend to physically cluster and separate from other conversations to be better understood and to enable listening, as depicted in Fig. 1. More interesting is the processing that we naturally do to effectively increase the signal to noise ratio (*S/N*). These largely take three forms based on binaural (two ears) listening [3]: interaural time differences (time of arrival or phase), interaural level differences (amplitude), and interaural decorrelation (coherence). These are similar to the diversity gain and orthogonal coding techniques used in wireless links to improve performance. Listeners with in only one active ear, for example a person using a hearing aid in one ear only, are severely disadvantaged in a high ambient noise environment since these binaural effects are not possible. Acoustic shadowing also occurs where the head blocks sounds and one ear can be favorably positioned. It has also been shown [3] that listeners, without knowing it, can also follow secondary conversations (data streams) simultaneous to the primary conversation with sudden awareness triggered by hearing their name, or certain key or taboo words. We also have extensive ability to correct for missing or mispronounced phonemes (differentiating sounds) and words, much like error correcting codes can overcome missing and incorrect bits.

The acoustics community has also developed quantitative tools to describe speech information that help or hinder understanding in the presence of noise. A basic approach is to divide the speech frequency range into a set of bands, rank these bands in terms of importance to speech comprehension, and track *S/N* in these bands for various types of noise signals. This allows both target and interfering sounds to be modeled in a repeatable way and provides insights into experiments designed to test comprehension versus noise. One result is that as the interfering signal fills the bands to similar levels as the target conversation, comprehension decreases. Thus, listeners will often have more difficulty understanding conversations if noise speakers have the same gender as the primary speaker, as similar speakers tend to have similar spectrums. This has parallels in the spoofing of wireless signals, where signals with similar but slightly corrupted modulation and frequency may cause more problems than simple higher amplitude white noise. Speech comprehension is also affected by sound level fluctuations (modulation), frequency hopping, time domain effects and clipping, spatial location of the sound sources, and more [3]. Whether these acoustic and speech insights can benefit wireless communications, MIMO processing techniques, and data coding protocols is an open question, but awareness of these results may be of interest.

Fig. 1. People having separate conversations tend to cluster and group as the ambient noise level rises.

This paper looks at only one small example of the above work on speech comprehension versus noise. A sound propagation model of the cocktail party effect can be developed as a simple approximate equation which gives a *S/N* value, namely, the acoustic energy density of the intended conversation over the reverberant energy density of the undesired background conversations. An electromagnetic equivalent to this acoustic *S/N* can be readily found using the same basic derivation. This *S/N* expression can form a starting point to think about the "coexistence" of wireless devices where a similar problem to the cocktail party exists, namely, how to maintain a desired link in the presence of multiple, allowable sources broadcasting in the same frequency band with similar spectral characteristics. We will first review the acoustic case in Section II followed by the electromagnetic equivalent in Section III. We then conclude with a brief discussion.

II. Multiple, Simultaneous Conversations in a Reverberant Room

The derivation of the basic equation describing the cocktail party effect may be found in [2, Chapter 6] and will be summarized here. Suppose there are K conversations occurring in a reverberant room. We assume one speaker per conversation at any given time with a speaking power P_t. Initially, we will let P_t be the same for each conversation cluster, but this can be readily generalized. Each conversation will involve some number of people, with the total number of people in the room being M ($M > 2K$). We will assume that each conversation cluster is separated from any other cluster by a distance greater than the radius of reverberation r_0; that is, the distance at which direct path energy density and reverberant paths energy density are equal. While speech is somewhat directional, we will assume for simplicity spherical spreading (i.e., isotropic radiation) for any given conversation. Then, the direct energy density w_d within a conversation is given by [2, eq. (1-12-1)]:

$$w_d = \frac{P_t / c}{4\pi r^2}, \tag{1}$$

where r is the distance between speaker and listener ($r < r_0$), and c is the sound speed. The reverberant energy density w_r due to the K conversations [2, eq. (6-2.13)] is simply K times the power per conversation P_t reduced by ¼ the room constant R_c (proportional to the room absorbing power):

$$w_r = \frac{4KP_t / c}{R_c} \quad , \tag{2}$$

where R_c is given by [2, eq. (6-2.13)]

$$R_c = \alpha S / (1-\alpha) \quad , \tag{3}$$

where S is the room surface area and α is the absorption coefficient for the wall surfaces, assumed uniform here over the whole room.

Combining these we have the total energy density w given by [2, eq. (6-3.9)]:

$$w = \frac{P_t}{c}\left(\frac{1}{4\pi r^2} + \frac{4K}{R_c}\right) \quad . \tag{4}$$

This expression can be used to give a criterion for a conversation to be understood by an isotropic listener, namely, that a conversation cluster "signal" energy density (the direct conversation plus the desired speaker's reverberant energy contribution) be greater than the "noisy" other K-1 cluster's reverberant energy density ($S/N > 1$):

$$S/N = \frac{\frac{1}{4\pi r^2} + \frac{4}{R_c}}{\frac{4(K-1)}{R_c}} \quad . \tag{5}$$

If we note that $r_0 = (R_c / 16\pi)^{1/2}$ [2, eq. (6-2.15)], then the above may be rewritten as [2, eq. (6.3-11)]:

$$S/N = \frac{(r_0 / r)^2 + 1}{K-1} \quad . \tag{6}$$

As is intuitively expected, we see that as the number of conversation clusters K increases, the *S/N* decreases. In addition, as K increases, then r must decrease to maintain the same S/N, implying that people in a conversation cluster must move closer together. This clearly sets a limit on K, and or M, as r may not be made arbitrarily small. Ignored here is the effect of adding people on the room constant R_c. In an acoustic room, the *S/N* ratio will drop too low and the room become too crowded well before R_c increases enough to make a difference. This may be different for RF reverberation chambers, however, which we consider next.

III. Multiple, Simultaneous RF Sources in a Reverberation Chamber

The above acoustic model has a direct analogy in a RF reverberation chamber. If we consider the average power density P_d (averaged over paddle stirring) in a reverberation chamber from K sources as consisting of a dominant direct path contribution (we will again initially assume isotropic radiation and reception) plus reverberant power (with all sources assumed to have equal strength), then we can write [6, eqs. (1-2)]

$$\langle P_d \rangle = P_t\left(\frac{1}{4\pi r^2} + \frac{K\lambda Q}{2\pi V}\right) , \tag{7}$$

where λ is the wavelength, Q is the chamber quality factor, and V the chamber volume. We note that Q is approximately equivalent to [4, eq. (5)], where

$$Q = 8\pi \frac{V}{\lambda \alpha S_c} \tag{8}$$

in the acoustic case (this approximation is based on normal incidence to the walls only as discussed in [7, Section 1.3], where α is the absorption coefficient and S_c is the surface area of the chamber. Inserting Q into (7) and noting that

$$\alpha S_c = R_c \tag{9}$$

for small α, (that is for highly reflective walls), we arrive at

$$\langle P_d \rangle = P_t \left(\frac{1}{4\pi r^2} + \frac{4K}{R_c} \right) , \tag{10}$$

which is equivalent to (4) for the reverberation chamber power density. Alternately, if we note that [6, eq. (3)] that

$$Q = \frac{V}{2\lambda r_e^2} , \tag{11}$$

then we see that (7) can also be rewritten as

$$\langle P_d \rangle = P_t \left(\frac{1}{4\pi r^2} + \frac{K}{4\pi r_e^2} \right) . \tag{12}$$

We can readily generalize this for unequal source powers to

$$\langle P_d \rangle = \left(\frac{P_1}{4\pi r^2} + \frac{\sum_{i=1}^{K} P_i}{4\pi r_e^2} \right) , \tag{13}$$

where P_1 represents the desired source to be received. The equivalent signal to noise ratio (6), that is, the desired direct path signal plus its desired signal reverberation over undesired signal reverberation is given by

$$S/N = \frac{P_1 \left((r_e / r)^2 + 1 \right)}{\sum_{i=1}^{K} P_i} . \tag{14}$$

Noting that for the undesired reverberant power only the sum of the powers matters and letting

$$P_{RT} = \sum_{i=1}^{K} P_i , \tag{15}$$

we arrive at

$$S/N = \frac{P_1 \left((r_e / r)^2 + 1 \right)}{P_{RT}} . \tag{16}$$

Finally, we may generalize this to include a directive intended source with directivity D_s and receiver D_r (these can be inserted in the direct power path in (13)) giving

$$S/N = \frac{P_1 (D_s D_r (r_c / r)^2 + 1)}{P_{RT}} . \tag{17}$$

As in the acoustic case, we see that as the total background reverberant power P_{RT} increases, then the distance from the desired source to the receiver will need to decrease to maintain an equivalent *S/N* or the directivity of the desired source D will need to increase. Alternately, the effective radius r_e , which is inversely proportional to Q, will need to increase.

The above discussion uses a simple ratio of powers to describe *S/N*. More interesting would be some communications metric, such as bit error rate (BER). We would not expect BER to follow a simple inverse relationship but there may be a S/N level as given by (17) that is representative of when BER reaches an unacceptable level for a given type of system or set of modulation characteristics. This would need to be explored with experimental data which is not considered here.

IV. Conclusion

We have reviewed some basic concepts in acoustics on how humans comprehend speech in the presence of background noise, the “cocktail party effect”, and how this relates to an electromagnetic equivalent. A simple power ratio is formed to represent signal to noise, or *S/N*. Future work could look at actual communication signal metrics such as BER, error vector magnitude (EVM), or other quantities to explore whether the *S/N* ratio given by (17) provides useful guidance on the performance of this metric. This may have application to coexistence standards where multiple sources and receivers are expected to operate in close proximity. More generally, the rich acoustics literature may provide additional insights into communications systems based on the ability of humans to process and filter signals in a variety of creative ways and how these can be translated to communications systems equivalents.

References

[1] C. Cherry, “Some experiments on the recognition of speech, with one and two ears,” Journal of the Acoustical Society of America, vol. 25, no. 5, pp. 975-979, 1953.

[2] A. Pierce, Acoustics: An Introduction to its Physical Principles and Applications, Acoustical Society of America, Ch. 6, 1989.

[3] A. Bronkhorst, “The cocktail-party problem revisited: early processing and selection of multi-talker speech,” Journal of Attention Perception & Psychophysics, vol. 77, no 5, pp. 1465-1487, April 2015

[4] P. Wilson. “Acoustic and electromagnetic reverberation chambers: similarities and differences,” in Proc. 2016 Asian-Pacific Intl. Symp. On Electromagn. Compat. (Shenzhen, China), May 2016.

[5] D. Wang, ”Deep learning reinvents teh hearing aid,” IEEE Spectrum, vol. 54, no. 3, pp. 32-37, March 2017.

[6] C. Holloway, D. Hill, J. Ladbury, and G. Koepke, “Requirements for an effective reverberation chamber: unloaded and loaded,” IEEE Trans. Electromagn. Compat., vol. 48, no. 1, pp. 187-194, Feb. 2006.

[7] D. Hill, Electromagnetics Fields in Cavities: Deterministic and Statistical Theories, IEEE Press, Wiley & Sons., 2009.