Auditory Informational Masking

Gerald Kidd Jr. and Christopher Conroy

Informational Masking (or Why Car Horns Shouldn't Talk)

Early in the career of one of the authors (Kidd), he worked as a research assistant on a project devoted to improving the emergency warning signals in airplanes (Tobias and Kidd, 1979). Tobias and Kidd investigated the use of auditory displays that spoke messages like "this way out" or "move toward exit," thereby providing verbal spatial indicators of the exit location in addition to the usual visual displays directing passengers. In such a context, even under the hectic and noisy scenario being emulated (using Gaussian noise), spoken messages could plausibly have beneficial effects on performance by helping passengers exit an airplane in an emergency.

Extrapolating from that situation, one could imagine that replacing the typical sound of a car horn with audible speech conveying meaningful messages might provide a similarly informative signal that could improve driving performance. Typical car horns currently produce sounds that serve as immediately identifiable warnings (although multiple similar-sounding horns honking simultaneously can be quite confusing) and that may provide indicators of the physical location of surrounding vehicles but otherwise convey little additional meaning. Why not increase the information a car horn conveys by having it produce messages comprising intelligible speech? Imagine if horns could say "passing on left" or "braking" or "get out of my way, you turkey!"

However appealing this might seem on the surface, equipping all cars with talking horns and creating a "cocktail party problem" (i.e., the difficult problem of conversing in situations like noisy, crowded rooms; see reviews in Middlebrooks et al., 2017) on the highway, strikes the current authors as a very bad idea for reasons that go to the heart of the phenomenon referred to as "informational masking" (Pollack, 1975). Informational masking means interference in understanding a signal that cannot be accounted for simply by the spectrotemporal overlap of competing sounds. This is in contrast to "energetic masking" that *can* be accounted for by the spectrotemporal overlap of competing sounds (cf. Kidd and Colburn, 2017).

Indeed, energetic masking is what usually comes to mind when one thinks of auditory masking. It occurs when one sound physically overpowers another and renders it inaudible. In contrast, informational masking is a broad characterization applied to many diverse listening situations. It typically refers to situations where the information necessary to solve the task is available, but for various reasons (e.g., perceptual, cognitive, linguistic), the listener cannot solve the task. In the talking car horn example, too many simultaneous sources of information that must be parsed and evaluated, especially when uncertainty and/or the acoustic similarity of the sources is high, can have a detrimental effect on communication leading to errors that could be catastrophic in certain real-world situations. We have learned through studies of masking that identifying and interpreting the messages from concurrent, independent sources of sound can be a challenging task even when there is no special

Figure 1. *Illustration of "talking car horns" (**message bubbles**) in traffic while our driver listens to a podcast (**rectangle**).*

https://doi.org/10.1121/AT.2023.19.1.29

information or urgency that is being conveyed. Processing more complicated messages can be taxing and may take resources away from other tasks, such as operating a moving vehicle. Some of the complexities of this imagined multitalker horn scenario are illustrated in cartoon fashion in **Figure 1**.

In **Figure 1**, our driver (listener) faces an extremely challenging task. They are trying to pay attention to a nontraffic acoustic source (the imaginary podcast *The Life of Bob*). Uncertainty is high because spoken messages could occur unexpectedly from different directions and each source must be parsed and interpreted as it occurs. Also, some proportion of the acoustic signals from different sources fall into the same frequency region(s) at the same time(s), obscuring whatever information the less intense sound conveys. This is also one way of describing energetic masking (e.g., a loud trash truck idles next to our driver's car and momentarily drowns out the podcast).

In a colloquial sense, the "information" conveyed by several talking car horns is much greater and more complex than if the sounds were simply car "honks," even though the number and level of the sound sources could be the same. All the sounds may be clearly audible and distinct (easily segregated) in either case, but the listener would have much greater difficulty navigating the sound field with the added burden of attending/ignoring/processing meaningful speech (i.e., understanding the messages of the talking horns and deciding on any actions that should be taken as a consequence). Such a complex task requires considerable mental processing that takes time and effort (e.g., Rennies et al., 2019).

There is also the fundamental problem that a sound source designated at one moment as an unwanted masker may suddenly become the desired target source requiring the refocusing of attention and, for speech, engaging the linguistic structures required to interpret the message. A basic function of audition is to constantly monitor the sound field even while focusing primarily on the current target source so that such source priority/designation shifts may occur. This means that we are always expending some of our available processing resources to evaluate (i.e., segregate, attend, remember, and, importantly, anticipate) the information from the various sources. The point is that multiple sources in a sound field producing complex messages concurrently pose a very significant challenge to a listener and tap into many levels of processing well beyond the acoustic overlap of the sounds or the associated competition for neural representation of the sounds in the auditory periphery. This is an example of auditory informational masking.

The Problem with Noise

Either unwanted car honks or the distracting verbal messages in the imaginary case illustrated in **Figure 1** could be considered "noise" if the goal of the listener was to only focus on the podcast and ignore everything else. More typically, some or all the sounds from the other cars would require some degree of processing, a portion of which might be obligatory for the task of safe driving.

The term noise has both scientific and everyday meanings. For that reason, unfortunately, there often is a great deal of imprecision in the way the term is used or how it is interpreted. Gaussian noise is a well-defined stochastic signal, whereas any type of unwanted or undesirable sound also qualifies as noise (e.g., American National Standard Institute, 2013). That definition depends, then, on the internal and changeable state of the listener. Thus, both Gaussian noise and the unwanted speech from another talker could be considered as noise, although either could also be the intended focus of attention under the appropriate circumstances. A classic illustration of this is the exchange between Sybil and Basil Fawlty in the episode "A Touch of Class" in the TV series *Fawlty Towers:* "Racket? That's Brahms! Brahms's *third* racket…" (BBC Productions, 1975), where "racket" is in the ear of the beholder.

In studies of auditory masking, the imprecise definition of noise often causes problems with the interpretation of experimental findings if all types of noise are considered the same because they qualify as "unwanted sounds." This means that the standard metric of signal-to-noise ratio may not be a reliable predictor of signal detection or recognition when considered across different types of noise.

Historically, perhaps because of the early development of the Gaussian noise generator and the predictable, repeatable masking such noise produces, the emphasis has been on energetic masking that has more or less served as the default masker for much of the perceptual and physiological work found in the literature. In fact, the difficulties and limitations caused by different types of unwanted sounds on the reception and understanding of a target sound vary dramatically and are based on many physical, perceptual, cognitive, and, for speech maskers, linguistic factors. For example, the speech from a masker talker may create less masking if it is spoken in a language that the listener does not understand than if it is spoken in the listener's primary language (e.g., Calandruccio et al., 2013) even if both signals are equally "unwanted." Or simply time reversing a speech masker (i.e., playing the waveform for each word in reverse so it is unintelligible but produces roughly the same energetic masking) can greatly reduce its effectiveness in masking an intelligible speech signal (Kidd et al., 2016). Indeed, it is the recognition of these important differences among various types of unwanted sounds that has, as much as anything, led to the subdivision of different classes of maskers into "energetic" and "informational" categories.

Additional Masking

A persistent problem in understanding the interference one speech source has on another for any task ranging from "simple" detection to comprehension is that the acoustic overlap between the two signals typically varies in different regions of the spectrum from moment to moment. This dynamic acoustic overlap and its counterpart in the internal neural representation in the listener makes the task of determining energetic masking for speech masking speech difficult. How, then, can we isolate and identify the different sources of masking in a speech mixture and, in particular, determine the influence of informational masking?

There have been many attempts over the years to find ways of separating energetic from informational masking in speech-on-speech situations (see Kidd and Colburn, 2017, for a review). **Figure 2** depicts a particularly successful approach, first reported by Brungart et al. (2006), that is called "ideal time-frequency segregation" (ITFS).

Figure 2 shows the spectrogram produced by a mixture of two different, concurrent words. One word was spoken by talker A while the other word was spoken by talker B. The relative intensities of the two words were plotted. Because these signals are known exactly, it is possible to calculate the relative intensities of each signal in each time-frequency (T-F) unit.

In **Figure 2**, the T-F units in which talker A are more intense than B are *red* (outlined with *boxes*) while the

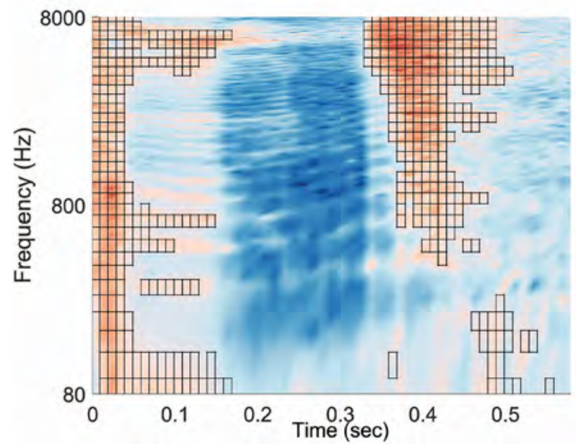


Figure 2. *A sound spectrogram of two words spoken concurrently by two different talkers. One word is in* ***red*** *and the other word is in* ***blue****. In time-frequency (T-F) regions where the two spectra overlap, the higher intensity signal is plotted. For the target (****red****) talker, the T-F units that are target dominated (****black rectangles****) are retained after ideal time-frequency segregation (ITFS) processing.*

T-F units dominated by talker B are *blue*. If talker A is the target, the units that are blue are considered to be *energetically masked* by talker B because there is more energy from talker B than talker A in those units. If we remove all the blue units by signal processing, what remains are the "glimpses" of the target speech that the user must rely on for intelligibility. For a highly informational masker like an intelligible talker, removing the masker-dominated T-F units can have an enormous effect because it eliminates the informational masking that is present.

What matters for intelligibility is the number of and/or the energy contained in the remaining glimpses (cf., Conroy et al., 2020), which depends on the level of the target relative to that of the masker (target-to-masker ratio [TMR]). It is possible to find a point of equal intelligibility (e.g., the proportion of energy that yields 50% correct performance), for target speech in noise or in speech maskers after ITFS processing. If one then takes the glimpsed stimuli derived from speech in noise and speech in speech that are equally intelligible and fills back in the full masker (i.e., restoring the unprocessed speech/noise mixture), the difference in intelligibility can be substantial (about 4 dB at threshold for noise compared with about 30 dB for speech; cf. Kidd et al., 2019), with much greater loss of intelligibility caused by

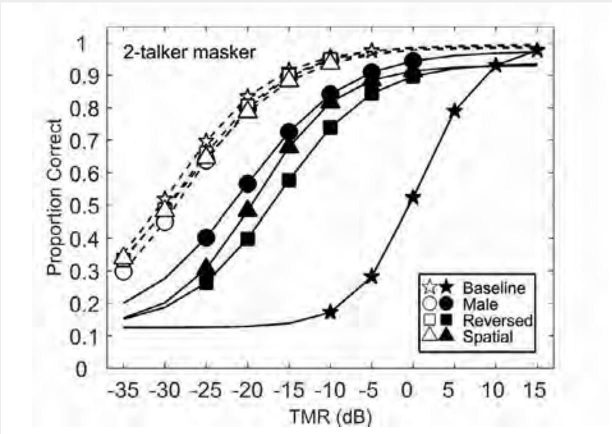


Figure 3. *Performance-level functions adapted from Kidd et al., (2016). The target is a female talker located at 0° azimuth.* ***Solid lines and symbols****, intelligibility for nonglimpsed speech;* ***dashed lines and open symbols****, for glimpsed speech (see text). The baseline condition (****stars****) comprised two same sex, colocated, natural speech maskers. The masker segregation conditions were male, time reversed, and spatially separated talkers. TMR, target-to-masker ratio. Reprinted from Kidd et al., (2016), with permission of Acoustical Society of America. © 2016, Acoustical Society of America.*

the speech masker than by the noise masker, even though the target information (i.e., the available glimpses of target speech) is about the same. This "additional masking" for the speech masker is informational masking. Examples of these sounds are provided in **Multimedia File 1** (see acousticstoday.org/kiddmedia). Some of the experimental findings supporting this point are shown in **Figure 3**.

Figure 3 shows the results of a speech-on-speech masking experiment in which several types of maskers were tested and ITFS processing was applied to each. The key point to focus on is the set of four psychometric functions on the left side of the graph (*dashed lines*). These psychometric functions show the intelligibility of the glimpsed target speech for each masker after removing the masker-dominated T-F units (i.e., after removing the informational masking) and are all about the same. However, the corresponding nonglimpsed maskers (**Figure 3**, *solid symbols*) produce markedly different amounts of informational masking depending on the particular target segregation cue available to the listener. This may be seen by comparing the TMR distance for glimpsed and nonglimpsed functions for the same masker types.

The take-away message here is that performance in speech-on-speech mixtures; specifically, those in which uncertainty and/or source similarity is high, is often dominated by informational rather than energetic masking. In realistic listening situations, many source segregation cues are available to help cause a release from informational masking. The ability to apply these various cues, which varies markedly across listeners, depends heavily on *context* and a priori *knowledge*.

Informational Masking and Detection Threshold

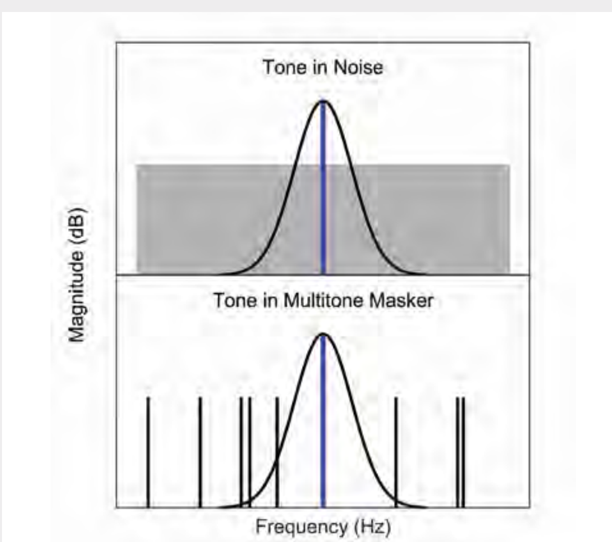
In the speech-on-speech masking experiment discussed in **Additional Masking** where all the signals are equal in level (i.e., 0 dB TMR equals a −3 dB signal-to-noise ratio for two masker talkers), there is enough information remaining after accounting for energetic masking (by ITFS) that the speech of the target talker would be nearly perfectly intelligible (see percent correct for glimpsed functions at 0 dB TMR in **Figure 3**). The interference the masker talker causes on the intelligibility of the target talker, then, is informational masking and is not due to a lack of audibility of the target speech. Thus, we say that the speech of both talkers is at *suprathreshold* levels. This raises the question, though, of what exactly is meant by detection threshold under informational masking?

A common theme in the interpretation of a masked detection threshold is that there is at least a rough correspondence between the physiological representation of the target signal and the behavioral performance on a psychophysical task. Framing this as a signal detection problem, the idea is that whatever distribution of physiological activity is relevant for solving the task, adding the signal shifts the distribution along the decision axis, resulting in better detection/discrimination performance (i.e., performance specified as the index of detectability [d′] improves as the distributions are separated; cf. Green and Swets, 1966). However, informational masking means that signals that presumably *should be* detectable based on the relevant underlying physiological distributions *are not*, as inferred from the behavioral performance of the observer. It should be noted, however, that demonstrations of the robust nature of the distributions of relevant physiological quantities for a target signal under informational masking conditions are difficult to obtain and the direct evidence for the relevant comparisons is limited.

Figure 4 illustrates how the energy from a masker falling in a hypothetical auditory filter measured psychophysically at a detection threshold for a pure-tone target is much less under informational masking than under energetic masking. Although the plots are illustrative and not drawn to scale, the evidence from the multitone masking literature (e.g., Durlach et al., 2005) suggests that, in some cases, detection thresholds may be more than 20 dB higher for equivalent masker energy in the signal's "critical band" under informational masking than under energetic masking (see Kidd et al., 2008, for a review).

For both cases, the spectra shown are for a single random sample of the masker; in an actual experiment, the noise varies from sample to sample as does the draw of tones in the multitone masker. An ideal observer (i.e., a model of performance in which decisions are based on selecting the most likely options; cf. Green and Swets, 1966) operating on the distribution of peripheral neural activity presumably would outperform the human observer considerably more for informational than for energetic masking conditions (e.g., Durlach et al., 2003). Thus, a paradox in defining informational masking for a detection threshold (and, by extension, some other tasks) is that the robustness of the neural representation of the target in the auditory periphery, yielding an equivalent psychophysical performance, presumably is much different under energetic and informational masking.

Figure 4. *A schematic illustration of a tone in noise detection task for two types of "noise."* ***Top:*** *the target tone (**blue line**) is masked by a broadband Gaussian noise (energetic masker; **gray rectangle**).* ***Bottom:*** *the target tone is masked by a random frequency (**black lines**) multitone masker (informational masker). A "critical band" filter is shown centered on the target frequency (**bell-shaped black curve** centered on **vertical blue line**).*



Suprathreshold Masking and the Cocktail Party Problem

Although informational masking has been shown to affect performance on many tasks ranging from detection through various types of suprathreshold discrimination and nonspeech pattern identification, it is the influence of informational masking on speech understanding that is most familiar and of the broadest general interest. Among the various factors that contribute to speech communication in multitalker listening situations, binaural analysis has perhaps received the greatest attention due in large part to the common takeaway message of the importance of spatial hearing from Cherry's (1953) seminal article defining the "cocktail party problem." As discussed in Kidd et al., (2008), Cherry (1953) also identified several other important factors in solving the cocktail party problem, notably source or message transition probabilities and presumably the understanding/exploitation of those probabilities in speech communication.

The preponderance of work on the benefits of binaural hearing for speech reception in noise has, in fact, examined masking conditions that were high in energetic masking or that did not attempt to separate energetic from informational factors. Historically, the "masking level difference" (MLD; Hirsh, 1948) and the "speech intelligibility level difference" (SILD; Licklider, 1948) for detecting a tone in Gaussian noise and for recognizing speech in Gaussian noise, respectively, have been advanced as compelling evidence for the important role of binaural analysis in improving speech recognition in noisy listening environments (cf. Green and Yost, 1975).

There is, however, an important distinction to be made between detecting or identifying speech in Gaussian noise and the same tasks for a speech signal under competition from concurrent talkers. In the former, high-energetic masking case, the phenomenon is an at- or near-threshold process limited in effect by a narrow dynamic range from chance-to-perfect performance, whereas in the latter, high-informational masking case,

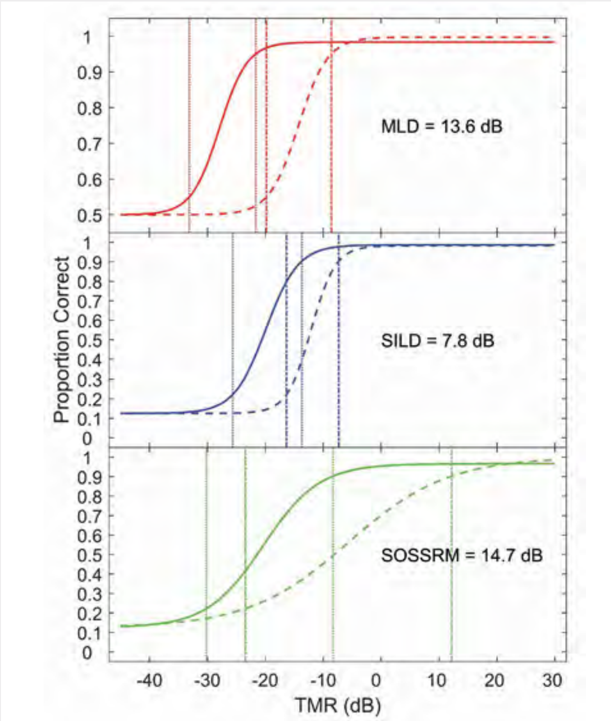


Figure 5. *Psychometric functions for three tasks.* ***Top:*** *homophasic (target and masker same in both ears;* ***dashed line****) and antiphasic (target π radians out of phase masker in phase) performance for detecting a tone in Gaussian noise (masking level difference [MLD];* ***solid line****).* ***Center:*** *same conditions for the task of speech intelligibility (speech intelligibility level difference [SILD]).* ***Bottom:*** *colocated (****dashed line****) and spatially separated (****solid line****) performance on a speech-on-speech masking task (SOSSRM).* ***Vertical lines:*** *points indicating 10% to 90% of the range of each function. The improvement in performance, release from masking, computed roughly at the middle of each function is indicated.*

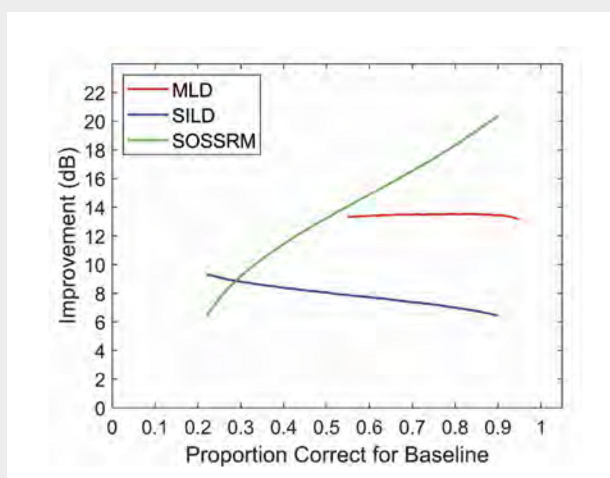
the phenomenon may be considered suprathreshold in nature and extends over a much wider range of levels above threshold. **Figure 5** illustrates this point.

Figure 5 shows the psychometric functions that illustrate the binaural advantages for three headphone-based tasks. **Figure 5**, *top and center*, demonstrates how binaural analysis provides a release from energetic masking (Gaussian noise masker) for detecting a pure tone (the MLD; **Figure 5**, *top*) and for speech intelligibility (the SILD; *center*). **Figure 5**, *bottom*, shows the release from informational masking through the spatial separation of sources (i.e., creating binaural differences at the ears via head-related transfer functions) for speech masking other speech (SOSSRM). **Figure 5**, *vertical lines*, indicate the *range* computed from 10% to 90% of each psychometric function.

In all cases, there is a significant separation between conditions where the waveforms are the same at the two ears ("diotic"; **Figure 5**, *dashed lines*) versus different at the ears ("dichotic"; **Figure 5**, *solid lines*). The improvement in performance, release from masking, computed roughly at the middle of each function is indicated. The slopes of the functions provide some indication of the mechanisms underlying performance. For the MLD, the slopes are parallel for the diotic and dichotic presentations, with function ranges of about 10 dB in each case. This is consistent with many past studies of the MLD (cf., Kidd et al., 1995). The SILDs are slightly less than the MLDs, and the slope of the dichotic function is shallower.

The conclusion is that, for these energetic maskers, the range between detecting a speech signal and fully understanding it is about 10-15 dB. Thus, the SILD operates near masked detection threshold. The functions for the speech maskers (SOSSRM) are much shallower than those for the SILD and extend over a broader range. The very shallow slope for the colocated condition reflects

Figure 6. *Release from masking for tone-in-noise detection (MLD;* ***red****), speech-intelligibility-in-noise (SILD;* ***blue****) and speech-on-speech masking (SOSSRM;* ***green****) calculated from the data plotted in* ***Figure 5****. The functions show release from masking as a function of the percent correct point on the reference (homophasic or colocated) psychometric function.*



the performance of some subjects who could segregate the sources by level and attend to the lower level talker (i.e., the target talker is intelligible at negative TMRs; cf. Byrne et al., 2022). These trends are shown in **Figure 6**.

When engaged in natural conversation, talkers typically raise the intensity of their voices to achieve high levels of intelligibility, usually requiring positive TMR values (cf. Weisser et al., 2021). It may be seen from **Figure 6** that the benefit of binaural/spatial cues increases as the intelligibility of the message is raised to comfortable conversational values. Because the long-term average speech spectrum peaks below 1 kHz and falls off above that value, the SILD is a mixture of the contributions of unmasking at different frequency regions at different TMRs. This means that the near-threshold binaural advantage for the SILD can be distributed over a wider range than occurs in any single frequency band (see discussion in Kidd et al., 1995). As energetic masking decreases (i.e., as the TMR increases), the relative influence of informational masking increases, as illustrated in **Figure 6** by the magnitude of the release from masking produced by the spatial separation of maskers. In other words, informational masking is primarily a suprathreshold phenomenon that dominates speech-on-speech masking across the range of TMRs typical of everyday speech communication.

Conclusions

Energetic masking is most likely a minor factor in cocktail party communication situations, with the greatest effect occurring within a few decibels of the detection threshold in any given frequency region. This conclusion is based on a consideration of the relative levels at which typical conversation takes place and an analysis of the glimpsed information available at those levels. However, it is probable that energetic masking interacts with informational masking in such situations to increase the communication difficulty (e.g., Best et al., 2020). Considering informational masking as what remains after accounting for energetic masking is a very broad definition and does not provide for the various, independent (to some degree) underlying physiological mechanisms and the subtleties of linguistic processing, especially under speech masking conditions.

Because of the high rate of word confusions found in some speech-masking tasks (e.g., Kidd et al., 2016) it is tempting to attribute informational masking in multitalker mixtures exclusively to misdirected attention. However, this too seems to be an oversimplification and difficult to support, in part because of the imprecision of the term “attention” (cf. Watson, 2005) as well as the existence of various other phenomena such as “linguistic informational masking” (e.g., Mepham et al., 2022), which argue against such a narrow interpretation. In the cocktail party problem, a listener must process the speech signal that has just arrived, drawing on memory and previously stored linguistic information while concurrently taking in new/ongoing information (see review in Kidd and Colburn, 2017). Uncertainty about the target speech along any of several dimensions undoubtedly imposes delays or even errors in accessing such information that could interfere with the processing of newly arriving sounds causing informational masking. There is also the very complex problem of actively disregarding unwanted sound sources while paradoxically monitoring them to a sufficient degree that the focus of attention may be redirected to them if circumstances warrant. Perhaps as important, though, is the ability of listeners to use recently arrived sounds to anticipate impending events; this may include leveraging expectation based on syntactic (e.g., Kidd et al., 2014) and semantic (e.g., Brouwer et al., 2012) probabilities. Although we have not discussed these issues in any detail in this article, both prediction and environmental monitoring are likely important in the context of informational masking.

Acknowledgments

Portions of this work were supported by the National Institute on Deafness and Other Communication Diseases, National Institutes of Health Grants R01-DC-004545 and R01-DC-013286. We appreciate the assistance provided by Christine Mason with the preparation of the manuscript and the contributions of our collaborators on several of the articles described herein.

References

American National Standard Institute (2013). *American National Standard Acoustical Terminology: Noise*. Acoustical Society of America, Melville, NY. Available at https://asastandards.org/terms/noise/.

Best, V., Conroy, C. and Kidd, G., Jr. (2020). Can background noise increase the informational masking in a speech mixture? *The Journal of the Acoustical Society of America Express Letters* 147, EL144-EL150. https://doi.org/10.1121/10.0000719.

Brouwer, S., Van Engen, K., Calandruccio, L., and Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content. *The Journal of the Acoustical Society of America* 131, 1449-1464. https://doi.org/10.1121/1.3675943.

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America* 120, 4007-4018. https://doi.org/10.1121/1.2363929.

Byrne, A. J., Conroy, C., and Kidd, G., Jr. (2022). The effects of uncertainty in level on speech-on-speech masking. *Trends in Hearing* 26, 1-16. https://doi.org/10.1121/1.1408946.

Calandruccio, L., Brouwer, S., Van Engen, K., Dhar, S., and Bradlow, A. R. (2013). Masking release due to linguistic and phonetic dissimilarity between the target and masker speech. *The American Journal of Audiology* 22, 157-164. https://doi.org/10.1044/1059-0889(2013/12-0072).

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *The Journal of the Acoustical Society of America* 25, 975-979. https://doi.org/10.1121/1.1907229.

Conroy, C., Best, V., Jennings, T., and Kidd, G., Jr. (2020). The importance of processing resolution in "ideal time-frequency segregation" of masked speech and the implications for predicting speech intelligibility. *The Journal of the Acoustical Society of America* 147, 1648-1660. https://doi.org/10.1121/10.0000893.

Durlach, N. I., Mason, C. R., Gallun, F. J., Shinn-Cunningham, B., Colburn, H. S., and Kidd, G., Jr, (2005). Informational masking for simultaneous nonspeech stimuli: Psychometric functions for fixed and randomly mixed maskers. T*he Journal of the Acoustical Society of America* 118, 2482-2497. https://doi.org/10.1121/1.2032748.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (2003). Note on informational masking (L). *The Journal of the Acoustical Society of America* 113(6), 2984-2987. https://doi.org/10.1121/1.1570435.

Green, D. M., and Swets, J. (1966). *Signal Detection Theory and Psychophysics*. John Wiley, New York, NY.

Green, D. M., and Yost, W. A. (1975). Binaural analysis. In Keidel, W. D., and Neff, W. D. (Eds.), *Handbook of Sensory Physiology: Auditory System*. Springer-Verlag, Berlin, Germany, pp. 461-480.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *The Journal of the Acoustical Society of America* 20, 536-544. https://doi.org/10.1121/1.1916992.

Kidd, G., Jr., and Colburn, H. S. (2017). Informational masking in speech recognition. In Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. R. (Eds.), *The Auditory System at the Cocktail Party*. Springer Nature, New York, NY, pp. 75-109.

Kidd, G., Jr., Mason, C. R., and Best, V. (2014). The role of syntax in maintaining the integrity of streams of speech. *The Journal of the Acoustical Society of America* 135, 766-777. https://doi.org/10.1121/1.4861354.

Kidd, G., Jr., Mason, C. R., and Rohtla, T. L. (1995). Binaural advantage for sound pattern identification. *The Journal of the Acoustical Society of America* 98, 1977-1986. https://doi.org/10.1121/1.414459.

Kidd, G., Jr., Mason, C. R., Best, V., Roverud, E., Swaminathan, J., Jennings, T., Clayton, K., and Colburn, H. S. (2019). Determining the energetic and informational components of speech-on-speech masking in listeners with sensorineural hearing loss. *The Journal of the Acoustical Society of America* 145, 440-457. https://doi.org/10.1121/1.5087555.

Kidd, G., Jr., Mason, C. R., Richards, V. M., Gallun, F. J., and Durlach, N. I. (2008). Informational masking. In Yost, W. A., Popper, A. N., and Fay, R. R. (Eds.), *Auditory Perception of Sound Sources*. Springer Science+Business Media, LLC, New York, NY, pp. 143-190.

Kidd, G., Jr., Mason, C. R., Swaminathan, J., Roverud, E., Clayton, K. K., and Best, V. (2016). Determining the energetic and informational components of speech-on-speech masking. *The Journal of the Acoustical Society of America* 140, 132-144. https://doi.org/10.1121/1.4954748.

Licklider, J. C. R. (1948). The influence of interaural phase relations on the masking of speech by white noise. *The Journal of the Acoustical Society of America* 20, 150-159. https://doi.org/10.1121/1.1906358.

Mepham, A., Bi, Y., and Mattys, S. L. (2022). The time-course of linguistic interference during native and non-native speech-in-speech listening. *The Journal of the Acoustical Society of America* 152, 954-969. https://doi.org/10.1121/10.0013417.

Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. R. (Eds.). (2017). *The Auditory System at the Cocktail Party*. Springer Nature, New York, NY.

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America* 57(Suppl. 1), S5. https://doi.org/10.1121/1.1995329.

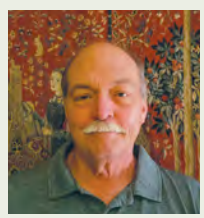
Rennies, J., Best, V., Roverud, E., and Kidd, G., Jr. (2019). Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and listening effort. *Trends in Hearing* 23, 1-21. https://doi.org/10.1177/2331216519854597.

Tobias, J. V., and Kidd, G., Jr. (1979). *Acoustic Signals for Emergency Evacuations*. Document #FAAAM795, Office of Aviation Medicine, Federal Aviation Administration, Department of Transportation, Washington, DC.

Watson, C. S. (2005). Some comments on informational masking. *Acta Acustica United with Acustica* 91, 502-512.

Weisser, A., Miles, K., Richardson, M. J., and Buchholz, J. M. (2021). Conversational distance adaptation in noise and its effect on signal-to-noise ratio in realistic listening environments. *The Journal of the Acoustical Society of America* 149, 2896-2907. https://doi.org/10.1121/10.0004774.

Contact Information



Gerald Kidd Jr. gkidd@bu.edu

Department of Speech, Language and Hearing Sciences
Boston University
635 Commonwealth Avenue
Boston, Massachusetts 02215, USA

Christopher Conroy
cconroy@sunyopt.edu

Department of Biological and Vision Sciences
State University of New York (SUNY)
College of Optometry
33 West 42 Street
New York, New York 10036, USA

For author bios, please go to acousticstoday.org/bios-19-1-3