Acoust. Sci. & Tech. **35**, 1 (2014)

INVITED REVIEW

The psychoacoustics of the irrelevant sound effect

Wolfgang Ellermeier[1,*] and Karin Zimmer[2,†]

[1]*Institut für Psychologie, Technische Universität Darmstadt, Alexanderstr. 10, D–64283 Darmstadt, Germany*
[2]*Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Schlossstr. 29, D–60486 Frankfurt, Germany*

Abstract: The decrement in memory performance observed while listeners are being exposed to acoustically structured stimuli is called the *irrelevant sound effect* (ISE). The present review summarizes the research identifying physical features of the irrelevant background that reliably induce performance decrements. It shows that speech, or speech analogues, produce the largest effects by far, suggesting that speech-specific features may contribute to auditory distraction. When an attempt is made to isolate psychoacoustical parameters contributing to the effect, it turns out that noticeable spectral change over time is a necessary condition to observe an ISE, while level change by itself is not. New empirical evidence is presented determining the rate of frequency modulation at which maximal effects are obtained. Results of a further study employing noise-vocoded speech show the importance of spectral detail in producing an ISE. At present, the wealth of empirical findings on the effects of irrelevant sound is not well accounted for by the available theoretical models. Cognitive models make only qualitative predictions, and psychoacoustical models (e.g., those based on fluctuation strength or the speech transmission index) account for subsets of the available data, but have thus far failed to capture the combined effects of temporal structure and spectral change in generating the interference.

Keywords: Irrelevant speech effect, Irrelevant sound effect, Auditory distraction, Noise effects, Fluctuation strength

PACS number: 43.50.Qp, 43.66.Ba [doi:10.1250/ast.35.10]

1. THE IRRELEVANT SPEECH EFFECT

The irrelevant speech effect (ISE) is a phenomenon originally observed in the psychological study of short-term memory [1,2]. It refers to the observation that the memorization of visually presented material, e.g., lists of letters or digits, is reliably and substantially impaired by the presence of background speech. This holds, even though the latter is entirely ‘irrelevant’ to the task at hand, typically presented at moderate levels, and participants are told to ignore it.

Figure 1 provides an example and illustrates the distribution and magnitude of effects typically obtained (Data from [3] re-analyzed). Subjects had to memorize random permutations of the digits 1 through 9 visually presented at a rate of 1/s while (a) overhearing a lecture in an unfamiliar language (here, Japanese), or (b) performing the same task in silence. Individual measures of ISE were obtained by subtracting the error rate with background speech from that obtained in silence, and referencing the change in performance to the baseline:

$$RDER = \frac{(E_{\mathrm{SND}} - E_{\mathrm{CONT}})}{E_{\mathrm{CONT}}} \quad (1)$$

where *RDER* is the *Relative Difference in Error Rate* (a measure that will be useful later in this review to compare the outcomes of different studies), E_{SND} is the number of errors (or, alternatively, error rate) in the presence of the sound studied (in this case, speech), and E_{CONT} is the one in a control condition (typically silence).

As is obvious from Fig. 1, irrelevant speech produces a decrement in short-term memory performance in most participants, while only few appear unaffected or do even better with background speech (negative scores); on average, by the measure defined in Eq. (1), the increase in error rate is approximately 50%.

Since it was demonstrated early on that speech, but not stationary broadband noise (e.g., [2,3]), produced the effect, and that it occured in memory (e.g., during retention, and not during encoding [4]), the ISE became essential in substantiating the phonological loop compo-

*e-mail: ellermeier@psychologie.tu-darmstadt.de
†e-mail: zimmer@dipf.de

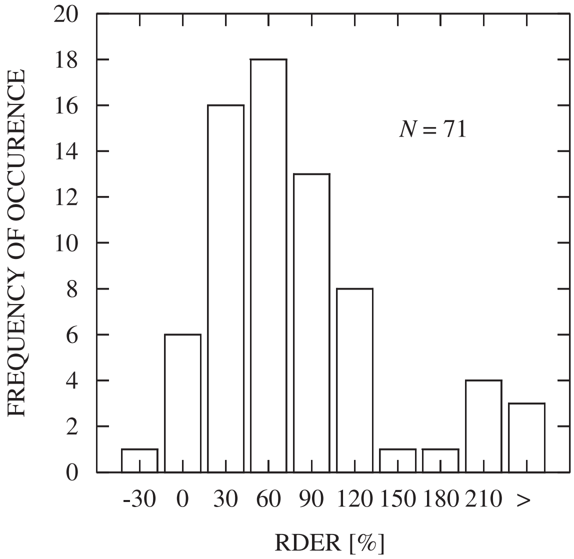

Fig. 1 Distribution of individual 'irrelevant speech effects,' based on data reported in [3]. The relative difference in error rate (*RDER*) due to irrelevant background speech, referenced to performance in silence (see Eq. (1)), was computed individually for each of 71 listeners. The rightmost bar refers to changes exceeding 240%.

nent of Baddeley's [5,6] modular working memory model (see Sect. 3.1).

When it became obvious that the effect was not restricted to speech, but occured with other acoustically structured, time-varying irrelevant material as well (see Sect. 2 and Table 1), it was re-labelled *irrelevant sound effect*, without having to change the acronym (*ISE*). As the vast body of research on the irrelevant sound effect has not been comprehensively reviewed since two nearly simultaneous papers by Jones and collaborators [7,8] appeared in 2001, another look at the effect and its varying theoretical accounts seems warranted, particularly, because (a) new empirical findings addressing the role of psychoacoustics in the ISE have emerged, and (b) several attempts to arrive at a unifying explanation of the effects observed have been proposed.

2. EFFECTIVE FEATURES OF THE IRRELEVANT SOUND

The focus of the present review is to explore whether there is enough evidence to assert which features of the acoustic background give rise to the detrimental effects on memory performance observed in the ISE paradigm. Therefore, it seems appropriate to identify a few robust, and numerously replicated findings about the ISE. These are:

(1) Stationary (e.g., unmodulated noise), or non-changing sounds (like repeated presentation of the same syllable) do not affect performance compared to quiet (e.g., [9]).
(2) The overall playback level does not seem to affect the magnitude of the ISE (e.g., [10,11]).
(3) The effect, as measured via Eq. (1), does not habituate. While, over a course of trials or sessions, memory performance improves, that happens to the same extent in irrelevant-sound and control conditions, leaving the crucial difference measure (the ISE) unchanged (see [11,12]).
(4) The semantics of the irrelevant speech do not seem to be important (e.g., [13]). Memory performance is equally affected whether participants hear an unfamiliar language or their own language in the background. Any residual effects of meaning are rather small (see [14–16]).

In our view, these key findings may be interpreted as favouring a psychoacoustical perspective on the ISE, encouraging efforts to identify the sound features producing it.

The next step then is to inspect the vast amount of studies in detail that have parametrically varied aspects of the irrelevant sound to identify the crucial, performance-relevant features. Table 1 lists a representative selection of these investigations, focusing on those that had scientific priority in studying a given parameter, but taking care that work from different laboratories is included as well. The table summarizes which physical features of the background signal have been varied, and states what effect sizes have been obtained. As an alternative measure of the ISE, and for easier comparison with other overviews [17,18] the *Absolute Difference in Error Rates*, *ADER* (where ER_{SND} and ER_{CONT} are the respective error rates in the irrelevant-sound and control conditions)

$$ADER = ER_{\mathrm{SND}} - ER_{\mathrm{CONT}} \qquad (2)$$

has been calculated as well, see the penultimate column of Table 1.

The first two sections of Table 1 show the large effect sizes obtained with free-running speech or sequences of different spoken words, and further demonstrate that almost equal effects may be obtained with music constituting the irrelevant background, suggesting that signals of sufficient complexity (i.e., spectro-temporal change) will yield maximal effects. The third section of Table 1 summarizes studies using speech signals played backwards, thus maintaining the spectral composition (and amount of spectral change) while reversing the temporal envelope. The fact that impairment is almost as large as with forward speech suggests that the acoustical features of speech are responsible for obtaining large ISEs, but that their meaning or phonetic identity is not crucial.

Table 1 Studies investigating selected parameters of the irrelevant sound.

Study	Source	(Psychoacoustic) parameter varied	Effect size	
			ADER [%]	RDER [%]
SPEECH				
[2] Salamé & Baddeley (1982)	Exp. 1	nonsense words	8.8	54.7
[19] Tremblay *et al.* (2000)	Exp. 1	words	29.0	96.6
[14] Buchner, Rothermund *et al.* (2004)	Exp. 1	words	14.0	32.5
[3] Ellermeier & Zimmer (1997)	Fig. 1	unfamiliar language	13.3	48.0
[20] Jones, Alford *et al.* (2000)[a]	Exp. 2	words	16.5	54.1
MUSIC				
[21] Salamé & Baddeley (1989)	Exp. 1	vocal music	14.2	38.8
[21] Salamé & Baddeley (1989)	Exp. 1	instrumental music	4.8	13.1
[22] Klatte & Hellbrück (1993)	Fig. 4	instrumental music	7.4	62.7
[23] Nittono (1997)	Tab. 1	instrumental music	5.8	29.3
[11] Ellermeier & Hellbrück (1998)	Exp. 1	(soft) music	7.6	37.3
BACKWARD SPEECH				
[13] Jones, Miles & Page (1990)	Exp. 5	backward speech	8.3	33.9
[16] LeCompte, Neely & Wilson (1997)	Exp. 4	reversed words	14.1	49.5
[24] Surprenant, Neath & Bireta (2007)	Exp. 3	backward speech	12.0	52.0
[25] Viswanathan *et al.* (2013)	Exp. 2	reversed sinewave speech	6.0	15.0
TONES AND FM				
[26] Jones & Macken (1993)	Exp. 1	varying tones	5.5	23.2
[15] Klatte, Kilcher & Hellbrück (1993)	Exp. 2	varying tones	5.2	25.2
[16] LeCompte, Neely & Wilson (1997)	Exp. 1	varying tones	4.0	11.8
[27] Jones, Macken & Murray (1993)	Exp. 1	interrupted FM tones	4.7	26.6
[20] Jones, Alford *et al.* (2000)[a]	Exp. 2	cello notes	7.0	29.2
[28] Hadlington *et al.* (2004)	Exp. 1b	varying tones	5.8	52.7
[29] Zimmer, Ghani & Ellermeier (2008)	Exp. 1	interrupted FM tones	4.5	17.4
LEVEL CHANGE AND AM				
[21] Salamé & Baddeley (1989)	Exp. 3	AM noise	1.2	3.9
[9] Jones *et al.* (1992)	Exp. 2b	AM noise	−1.1	−5.7
[15] Klatte, Kilcher & Hellbrück (1993)	Exp. 3	AM 'speechlike' noise	2.7	11.9
[30] Tremblay & Jones (1999)	Exp. 1	changing level	2.2	5.9

[a]Note: Since no silent baseline condition was available, the most degraded (noise-like) condition was used as reference.

The final section of Table 1 looks at the effect of amplitude change in the background signal, and unanimously finds it to fail to produce irrelevant-speech effects. Thus amplitude-modulated noise, or even noise shaped with a temporal envelope as is typical for speech (see [12] and Sect. 2.2), will not produce significant amounts of interference with the memory task.

The penultimate section of Table 1 focusses on a variety of manipulations the essence of which is varying pitch (via tone frequency or frequency modulation) in the background sound sequence. It appears that the effects of non-speech stimuli of this kind tend to yield smaller, but still significant effects on performance in ISE memory tasks, suggesting that pitch change is necessary, but may not be sufficient to generate large ISEs. It further appears that some kind of segmentation of the signals (discrete tones or interrupted FM glides) is required to trigger an ISE. That shall be further explored in Sect. 2.1.

2.1. Varying Frequency

The review presented in Table 1 suggests that pitch change between successive sound events in the background stream is crucial in order to produce an ISE. Few studies, however, have parametrically varied the amount of pitch change in the distracting signal. Therefore, in an unpublished study performed in our laboratory [31], we investigated both continuous and regularly interrupted sinusoidal frequency modulation (FM) while varying the FM rate between 0 (no modulation) and 50 Hz. Figure 2 shows a re-analysis of these data based on the performance of 24 participants.

It is evident that none of the frequency-modulated tones used boosts the error rate as much as speech does: While the latter increases the number of errors by some 50% (re Eq. (1)) on average, the relative increase in errors produced by the most harmful FM sound is just 22.5%. More interesting is the fact, that errors vary nonmonotoni-

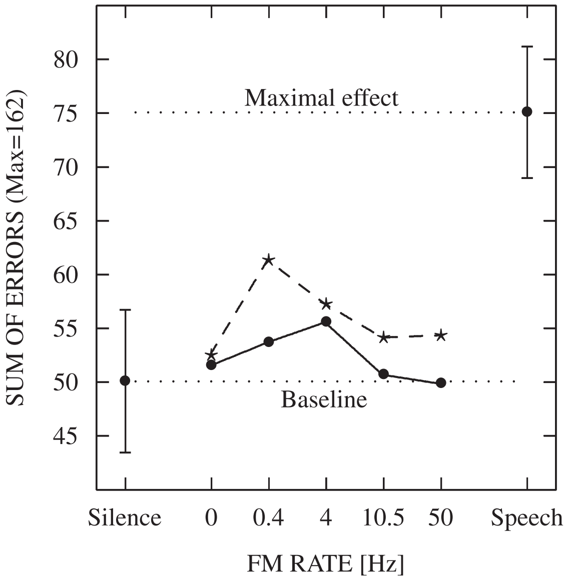

Fig. 2 Effect of the rate of frequency modulation on serial recall errors. Average sum of errors made by 24 subjects. The bracketing conditions 'Silence' and (unfamiliar) 'Speech' are plotted along with inter-individual standard errors. Asterisks refer to performance with regularly interrupted FM glides, filled circles connected by solid lines to non-interrupted sinusoidal FM.

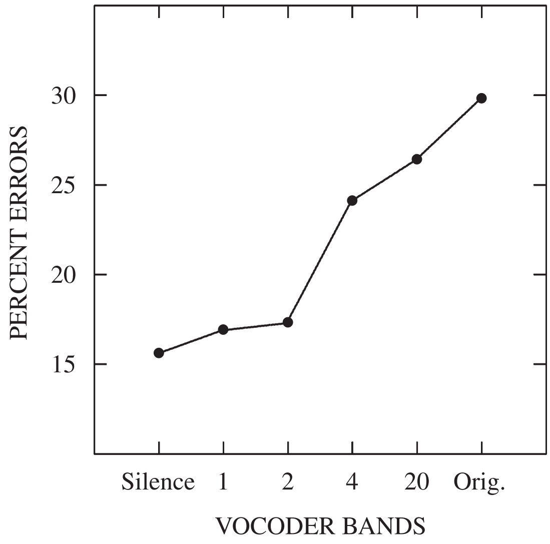

Fig. 3 Error rate as a function of the number of channels used to noise-vocode speech. Data (extracted from [33]) reflect the performance of 20 Japanese listeners exposed to irrelevant (German) speech. The rightmost data point marks the original, unprocessed utterances.

cally as a function of FM rate, both for continuous and interrupted FM tones. For the former, the present data suggest a maximum near 4 Hz, for the latter it is even lower. That is intriguing, since psychophysical scaling studies have found human perception of slow auditory fluctuations (i.e., the sensation of *fluctuation strength*) [32] to exhibit a maximum at a modulation rate of 4 Hz (see Sect. 3.2). What is further noteworthy, is that (in contrast to [27]) even non-interrupted FM appears to follow that pattern generating a small performance decrement (by some 11%) near that frequency.

2.2. Varying Spectral Detail

As it appears that reducing spectral change to variations of a single frequency trace (by varying the frequency of discrete tones or the FM rate of a modulated sinusoid) does not seem to trigger a maximal ISE, we have since taken the opposite route, by successively removing spectral detail from a complex speech signal [33]. That was accomplished by passing a corpus of sentences through a noise vocoder and varying the number of filters used to manipulate the spectral detail being rendered. Using 20, 4, 2, or just one such filter — their frequency boundaries being determined by earlier research on finding optimal subdivisions across languages [34] — a graded transition between the original, unprocessed speech, and pure, though speech-like, amplitude modulation of broadband noise (one band) was generated.

Figure 3 depicts a subset of the results obtained when 20 Japanese participants were exposed to German noise-vocoded speech while trying to memorize number sequences. Errors in serial recall were not affected (compared to work in silence) when the speech signal was passed through a single or just two vocoder bands; they appreciably increased, however, almost doubling the percentage of errors as the number of frequency bands was increased to 4 or 20, thus making more spectral detail available.

These results suggest that spectral change (not just temporal amplitude modulation) is crucial to obtain irrelevant-sound effects, and that, taking into account the weak effects obtained with pure or frequency-modulated tones (Sect. 2), it has to be fairly complex in order to produce maximal interference. In the following we will discuss, how the various theoretical models of the ISE deal with the effects of the psychoacoustical manipulations that were shown to be substantial.

3. MODELING THE ISE

Over the nearly 40 years of research on the ISE, several theoretical accounts of the effect have been proposed. Within the scope of this paper we will investigate whether they are consistent with the psychoacoustical implications of the studies reviewed.

3.1. Cognitive Models

The initial explanation of the irrelevant speech effect was part of a model of working memory proposed by Baddely and collaborators [5,6]. The model has a modular structure in that a visuo-spatial sketchpad and a phonological store are distinguished. It is in the latter that the to-be-remembered items, whether presented visually or acoustically, are supposed to be stored (and maintained via rehearsal) while the irrelevant speech (by its phonological nature) obtains obligatory access to it, thus competing for the same resources. It appears that this model does well in explaining the superiority of speech in producing interference, and particularly effects (for which there is little empirical evidence, though [35]) of phonological similarity between to-be-remembered material and distractors. The model is not explicit though, on which speech-like features are crucial to produce disruption, and thus, in its strictest form, does not account for non-speech ISEs.

Therefore, Jones and co-workers proposed a competing explanation according to which it is not the speech-like character, but the *changing-state* nature of the irrelevant sound that is responsible for the effect. Their *object-oriented episodic record (O-OER)* model [36] assumes that objects (the to-be-rememberd items as well as the elements of the irrelevant sound) entering episodic memory are joined by pointers that keep a record of their succession. When the background sound is of a steady-state nature (as with continous noise or the repetition of a single syllable) no such pointers (or new objects) have to be generated, the record is said to be self-referential [27]. When, by contrast, the irrelevant sound is of a changing-state nature, many such pointers will automatically keep track of the succession of sound elements. It is the two kinds of serial-order information about (a) the to-be-learned sequence of items and (b) the automatically registered succession of auditory events that interfere with each other in a single, and multimodal memory representation.

One of the strongest points in favour of this model is that it accounts for the effects of the type of task used in the ISE paradigm: Note that ISEs are strongest when the focal (memory) task requires keeping track of order information, as in the serial recall of previously presented items; and the ISE is greatly attenuated when other types of dependent measures are used (such as free recall, text comprehension, identifying a missing item, and the like) [7,37].

As the focus of this review is concerned: This model is also much more suited to account for effects due to the acoustical structure of the irrelevant sound, particularly when the interpretation of changing state(s) is not merely acoustical, but psychoacoustical, taking, for example, principles of auditory streaming, i.e., the formation of perceptual units, into account [38]. Thus the studies listed in Table 1 and many others may be seen as providing ample evidence for the power of the 'changing-state' heuristic. Nevertheless, it is not quite clear what constitutes a changing state (a pitch shift may qualify; a level-alteration not), and the reasoning is often at risk to become circular: If it produces an ISE, the signal must contain a changing state.

3.2. Psychoacoustic Models

With respect to the aim of this review, the drawback of the cognitive models discussed thus far is that they are neither quantitative, nor psychoacoustic in nature. These models are suited to summarize a body of research with respect to qualitative principles (e.g., by specifying which type of irrelevant sound will interfere most with which type of task), but they do not make quantitative predictions. Furthermore, they are not psychoacoustic in that they would predict a certain amount of disruption based on the auditory-perceptual features of the irrelevant sound.

An initial attempt was made by proposing the *speech transmission index* (STI) — an objective measure of the quality of speech and its potential intelligibility — as a unique predictor of the amount of interference in the ISE paradigm [17]. This measure does well in accounting for situations of partial masking or degraded transmission, but requires prior knowledge of the signal, and is not well suited to explain the non-speech ISEs observed in the literature.

More recently, a group of psychoacousticians has proposed a predictive model relating the ISE to an auditory sensation that has emerged as one of the basic dimensions by which sounds can be distinguished, i.e., their *fluctuation strength* [32]. Schlittmeier *et al.* [18] reviewed an impressive number of data sets ($N = 70$) from their own laboratories using a wide variety of irrelevant sounds (speech, masked speech, music, animal calls, etc.) and found that the magnitude of the ISE (as defined by Eq. (2)) could be predicted quite well by the auditory fluctuation strength (F) determined from the signals:

$$ADER = \frac{F}{0.68\,\mathrm{vacil}}\,7.5 \tag{3}$$

A simple linear regression of the absolute difference in error rates (ADER, i.e., the ISE) on fluctuation strength, referenced as specified in Eq. (3), and thereby cancelling the unit of measurement (vacil), accounted for 55% of the covariation between the two measures. That is quite impressive, given that only a single sound feature served as a predictor, but may not be telling the whole story as yet. Note that the psychoacoustical metric of fluctuation strength does not distinguish between amplitude (AM) and frequency (FM) fluctuations, thus implying that an amplitude-modulated noise with a modulation rate of 4 Hz should have a maximal effect, which, by the empirical evidence (see Table 1), it does not. Likewise, the effect of the number of channels in noise-vocoded speech (see Fig. 3 and [33]) is not likely to be predicted by the algorithm, since amplitude fluctuations are present irrespective of the number of channels. The authors note this shortcoming, and suggest to supplement the algorithm with some mechanism to tell whether successive sound elements are the same or different (the notion of changing state, see Sect. 3.1).

This idea was pursued in another study [39] that employed an adaptive masking scheme to reduce the temporal distinctiveness of successive speech segments, thereby attenuating the effects of irrelevant speech. While the success of this manipulation, a reduction in error rate by 9% (re Eq. (2)) compared to unaltered speech, was not well predicted by the STI, an ad-hoc algorithm quantifying the amount of spectral change occuring between successive segments, fared much better. The authors conclude that both spectral and temporal estimators (like the STI) should be used to predict the effects of irrelevant sound on performance.

4. CONCLUSIONS AND FUTURE WORK

The evidence reviewed suggests that there is a strong psychoacoustical basis for the irrelevant sound effect. Manipulation of the acoustics of the irrelevant sound has a systematical, and predictable influence on the magnitude of the ISE. As can be seen in Table 1 there appears to be — despite considerable variability — a continuum ranging from the small effects produced by pure amplitude or frequency variation over the medium effect sizes produced by backward speech or music to the maximal effects obtained with free-running, natural speech. From a psychoacoustical viewpoint, the basic hearing sensations related to amplitude and frequency variations, i.e., loudness, pitch, and fluctuation strength, have been thoroughly investigated. It appears less likely that other sensations typically studied in psychoacoustics (such as roughness, tonal content, and sharpness, see [32]), when varied within the irrelevant stream, might also produce ISEs, but this remains an open issue.

The empirical evidence mustered in this review suggests that the presence of pitch changes (see Fig. 2), when they occur at appropriate rates, is a necessary condition to obtain irrelevant sound effects. It is not sufficient, however, to produce maximal effects. These are only observed, when the irrelevant sound has rich spectral detail (see the effects of noise-vocoding in Fig. 3), and when that kind of detail is spectrally smeared, the ISE eventually vanishes.

That kind of evidence further suggests that the predictive algorithms presently suggested, though performing well for a limited set of conditions like partially masked irrelevant speech, will not account for the range of sound manipulations reviewed. Specifically, psychoacoustical fluctuation strength [18] or the speech transmission index (STI) [17] tend to respond to amplitude fluctuations that by themselves do not give rise to an ISE (see the 1-channel condition in Fig. 3). Furthermore, these models have difficulty explaining the unique disruptive potential of speech, compared to all other equally variable backgrounds (e.g., music or reversed speech).

The need to look into speech-like or speech-specific features in the irrelevant sound was emphasized by a recent study [25] that used variations of sinewave speech, a speech analogue in which the original speech signal is replaced by a series of time-varying sinusoids tracking formant centers, in the ISE paradigm. It was found that while sinewave speech disrupted performance as much as did natural speech, a condition in which the sinusoidal components representing two of three formants were time-reversed did not. Thus changing the dynamic relations between formants reduced the ISE (while supposedly maintaining the amount of changing state), suggesting that speech-specific properties of the background signal may be crucial for obtaining maximal effects. It appears that further research into the role of these features will have to be done and that using a single psychoacoustical dimension as a predictor may be insufficient to account for the complexity of the effects.

[The Japanese translation of this article was published in *the Journal of the Acoustical Society of Japan*, **69**, 638–646 (2013).]

REFERENCES

[1] H. A. Colle and H. Welsh, "Acoustic masking in primary memory," *J. Verb. Learn. Verb. Behav.*, **15**, 17–32 (1976).

[2] P. Salamé and A. Baddeley, "Disruption of short-term memory by unattended speech: Implications for the structure of working memory," *J. Verb. Learn. Verb. Behav.*, **21**, 150–164 (1982).

[3] W. Ellermeier and K. Zimmer, "Individual differences in susceptibility to the irrelevant speech effect," *J. Acoust. Soc. Am.*, **102**, 2191–2199 (1997).

[4] C. Miles, D. M. Jones and C. A. Madden, "Locus of the irrelevant speech effect in short-term memory," *J. Exp. Psychol. Learn. Mem. Cognit.*, **17**, 578–584 (1991).

[5] A. Baddeley, *Working Memory* (Oxford University Press, New York, 1986).

[6] A. Baddeley, "Working memory: Looking back and looking forward," *Nat. Rev. Neurosci.*, **4**, 829–839 (2003).

[7] S. P. Banbury, W. J. Macken, S. Tremblay and D. M. Jones, "Auditory distraction and short-term memory: Phenomena and practical implications," *Hum. Factors*, **43**, 12–29 (2001).

[8] R. Hughes and D. M. Jones, "The intrusiveness of sound: Laboratory findings and their implications for noise abatement," *Noise Health*, **4**, 55–74 (2001).

[9] D. M. Jones, C. A. Madden and C. Miles, "Privileged access by irrelevant speech to short-term memory: The role of changing state," *Q. J. Exp. Psychol.*, **44A**, 645–669 (1992).

[10] H. A. Colle, "Auditory encoding in visual short-term recall: Effects of noise intensity and spatial location," *J. Verb. Learn. Verb. Behav.*, **19**, 722–735 (1980).

[11] W. Ellermeier and J. Hellbrück, "Is level irrelevant in 'irrelevant speech'? Effects of loudness, signal-to-noise ratio, and binaural unmasking," *J. Exp. Psychol. Hum. Percept. Perform.*, **24**, 1406–1414 (1998).

[12] J. Hellbrück, S. Kuwano and S. Namba, "Irrelevant background speech and human performance: Is there long-term habituation?," *J. Acoust. Soc. Jpn.* (*E*), **17**, 239–247 (1996).
[13] D. M. Jones, C. Miles and J. Page, "Disruption of proofreading by irrelevant speech: Effects of attention, arousal, or memory?," *Appl. Cognit. Psychol.*, **4**, 89–108 (1980).
[14] A. Buchner, K. Rothermund, D. Wentura and B. Mehl, "Valence of distractor words increases the effects of irrelevant speech on serial recall," *Mem. Cognit.*, **32**, 722–731 (2004).
[15] M. Klatte, H. Kilcher and J. Hellbrück, "Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen," *Z. Exp. Psychol.*, **42**, 517–544 (1995).
[16] D. C. LeCompte, C. B. Neely and J. R. Wilson, "Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect," *J. Exp. Psychol. Learn. Mem. Cognit.*, **23**, 1–12 (1997).
[17] V. Hongisto, "A model predicting the effect of speech of varying intelligibility on work performance," *Indoor Air*, **15**, 458–468 (2005).
[18] S. J. Schlittmeier, T. Weißgerber, S. Kerber, H. Fastl and J. Hellbrück, "Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength," *Atten. Percept. Psychophys.*, **74**, 194–203 (2012).
[19] S. Tremblay, A. P. Nicholls, D. Alford and D. M. Jones, "The irrelevant sound effect: Does speech play a special role?," *J. Exp. Psychol. Learn. Mem. Cognit.*, **26**, 1750–1754 (2000).
[20] D. M. Jones, D. Alford, W. J. Macken, S. P. Banbury and S. Tremblay, "Interference from degraded auditory stimuli: Linear effects of changing state in the irrelevant sequence," *J. Acoust. Soc. Am.*, **108**, 1082–1088 (2000).
[21] P. Salamé and A. Baddeley, "Effects of background music on phonological short-term memory," *Q. J. Exp. Psychol.*, **21**, 107–122 (1989).
[22] M. Klatte and J. Hellbrück, "Der 'Irrelevant Speech Effekt': Wirkungen von Hintergrundschall auf das Arbeitsgedächtnis," *Z. Lärmbekämpfung*, **40**, 91–98 (1993).
[23] H. Nittono, "Background instrumental music and serial recall," *Percept. Mot. Skills*, **84**, 1307–1313 (1997).
[24] A. M. Surprenant, I. Neath and T. J. Bireta, "Changing state and the irrelevant sound effect," *Can. Acoust.*, **35**, 86–87 (2007).
[25] N. Viswanathan, J. Dorsi and S. George, "The role of speech-specific properties of the background in the irrelevant sound effect," *Q. J. Exp. Psychol.*, online preprint (2013).
[26] D. M. Jones and W. J. Macken, "Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory," *J. Exp. Psychol. Learn. Mem. Cognit.*, **19**, 369–381 (1993).
[27] D. M. Jones, W. J. Macken and A. C. Murray, "Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation," *Mem. Cognit.*, **21**, 318–328 (1993).
[28] L. Hadlington, R. M. Bridges and J. Darby, "Auditory location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear, or both ears," *Brain Cognit.*, **55**, 545–557 (2004).
[29] K. Zimmer, J. Ghani and W. Ellermeier, "The role of task interference and exposure duration in judging noise annoyance," *J. Sound Vib.*, **311**, 1039–1051 (2008).
[30] S. Tremblay and D. M. Jones, "Change of intensity fails to produce an irrelevant sound effect: Implications for the representation of unattended sound," *J. Exp. Psychol. Hum. Percept. Perform.*, **25**, 1005–1015 (1999).
[31] U. Wolski, *Experimentelle Untersuchung der Wirkung frequenzmodulierten Hintergrundschalls auf die Leistung in einer Gedächtnisaufgabe* [Experimental study of the effect of frequency modulated tones on performance in a memory task] (Unpublished diploma thesis, Universität Regensburg, 1995).
[32] H. Fastl and E. Zwicker, *Psychoacoustics — Facts and Models*, 3rd ed. (Springer, Berlin, 2007).
[33] W. Ellermeier, F. Kattner, K. Ueda, K. Doumoto and Y. Nakajima, "Effects of noise-vocoded speech in the irrelevant sound paradigm," *Fortschritte der Akustik — DAGA 2012*, pp. 336–337 (2012).
[34] K. Ueda and Y. Nakajima, "A consistent clustering of power fluctuations in British English, French, German, and Japanese," *Trans. Tech. Comm. Physiol. Psychol. Acoust. Acoust. Soc. Jpn.*, **38**, 771–776 (2008).
[35] D. C. LeCompte and D. M. Shaibe, "On the irrelevance of phonological similarity to the irrelevant speech effect," *Q. J. Exp. Psychol.*, **50A**, 100–118 (1997).
[36] D. M. Jones, P. Beaman and W. J. Macken, "The object-oriented episodic record model," in *Models of Short-term Memory*, S. E. Gathercole, Ed. (Psychology Press, Howe, England, 1996), pp. 209–237.
[37] D. C. LeCompte, "Extending the irrelevant speech effect beyond serial recall," *J. Exp. Psychol. Learn. Mem. Cognit.*, **20**, 1396–1408 (1994).
[38] W. J. Macken, S. Tremblay, R. J. Houghton, A. P. Nicholls and D. M. Jones, "Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory," *J. Exp. Psychol. Hum. Percept. Perform.*, **29**, 43–51 (2003).
[39] M. Park, A. Kohlrausch and A. van Leest, "Irrelevant speech under stationary and adaptive masking conditions," *J. Acoust. Soc. Am.*, **134**, 1970–1981 (2013).

Wolfgang Ellermeier received his Ph.D. in experimental psychology from the State University of New York at Stony Brook in 1988. Subsequently he was a postdoc at the universities of Würzburg and Regensburg, Germany. In 2001 he was appointed to head the Sound Quality Research Unit (SQRU) at the University of Aalborg, Denmark. Since 2007 he is a professor of Applied Cognitive Psychology at Technische Universität Darmstadt, Germany.

Karin Zimmer received her Masters and Ph.D. degrees in psychology from the University of Regensburg, Germany. After a postdoc in psychoacoustics and an associate professorship at the University of Aalborg, Denmark, she worked for the OECD in Paris on educational measurement. She is currently a researcher with the German Institute for International Pedagogical Research (DIPF) in Frankfurt, Germany, and coordinator of the Leibniz Education Research Network (LERN).