Algorithmic modeling of the irrelevant sound effect (ISE) by the hearing sensation fluctuation strength

Sabine J. Schlittmeier · Tobias Weißgerber ·
Stefan Kerber · Hugo Fastl · Jürgen Hellbrück

Abstract Background sounds, such as narration, music with prominent staccato passages, and office noise impair verbal short-term memory even when these sounds are irrelevant. This irrelevant sound effect (ISE) is evoked by so-called changing-state sounds that are characterized by a distinct temporal structure with varying successive auditory–perceptive tokens. However, because of the absence of an appropriate psychoacoustically based instrumental measure, the disturbing impact of a given speech or nonspeech sound could not be predicted until now, but necessitated behavioral testing. Our database for parametric modeling of the ISE included approximately 40 background sounds (e.g., speech, music, tone sequences, office noise, traffic noise) and corresponding performance data that was collected from 70 behavioral measurements of verbal short-term memory. The hearing sensation fluctuation strength was chosen to model the ISE and describes the percept of fluctuations when listening to slowly modulated sounds ($f_{mod} < 20$ Hz). On the basis of the fluctuation strength of background sounds, the algorithm estimated behavioral performance data in 63 of 70 cases within the interquartile ranges. In particular, all real-world sounds were modeled adequately, whereas the algorithm overestimated the (non-)disturbance impact of synthetic steady-state sounds that were constituted by a repeated vowel or tone. Implications of the algorithm's strengths and prediction errors are discussed.

Keywords Speech perception · Psychoacoustics · Short-term memory

The corresponding author provides a DVD with all background sounds included in the database upon request.

S. J. Schlittmeier · J. Hellbrück
Catholic University of Eichstaett-Ingolstadt,
Eichstaett, Germany

T. Weißgerber · S. Kerber · H. Fastl
AG Technische Akustik, MMK,
TU München, Germany

S. J. Schlittmeier (✉)
Work, Environmental and Health Psychology,
Catholic University of Eichstätt-Ingolstadt,
85071 Eichstätt, Germany
e-mail: Sabine.Schlittmeier@KU-Eichstaett.de

Introduction

Certain background sounds disturb verbal short-term memory. This so-called irrelevant sound effect (ISE) is an empirically robust phenomenon and occurs even when backgrounds sounds are irrelevant to the task and are ignored. A multitude of studies have explored the necessary sound characteristics for an ISE to be elicited (for a summary, see Hellbrück & Liebl, 2008). Nonetheless, no psychoacoustically based instrumental measurement procedure has been made available that can predict the detrimental impact of a given sound. Rather, the crucial sound criteria are still described in qualitative–descriptive terms, and behavioral experiments are indispensable in determining the disturbance impact of a given sound. As such, this article presents an algorithm that models the adverse effects of speech and nonspeech sounds on verbal short-term memory (i.e., the ISE) based on the hearing sensation fluctuation strength. Thus, a link between a psychoacoustical measure and cognitive performance is provided. To understand the description of the algorithm and its derivation, it is first necessary to consider the ISE and characteristic features of performance-reducing background sounds in detail.

Irrelevant sound effect

The standard task for measuring verbal short-term memory capacity, and thus for ISE exploration, is verbal serial recall. In the present study, unrelated verbal items (e.g., digits, consonants, words) were presented successively and had to be recalled afterwards in the exact presentation order. The irrelevant background sound is typically played until the entire recall task is completed. Colle and Welsh (1976) described a reduction in verbal short-term memory during irrelevant background speech for the first time. Since this initial demonstration, numerous behavioral studies have explored the characteristics of sounds that are necessary and sufficient to adversely affect cognitive performance. The ISE is defined as the performance decrement during a given background sound as compared with performance during silence. For an ISE to occur, the crucial factor is the inherent auditory–perceptive properties of the irrelevant sound. Research has suggested that short-term memory is reduced by irrelevant speech and nonspeech sounds that are characterized by changing-state features; that is, distinctive temporal–spectral variations allow for perceptual segmentation of an irrelevant sound while, at the same time from an auditory-perceptive view, successive perceptual tokens also change. Likewise, narration (e.g., Buchner, Irmen, & Erdfelder, 1996; Colle & Welsh, 1976; Salamé & Baddeley, 1986, 1987), irrelevant speech consisting of different consonants (e.g., Jones, Madden, & Miles, 1992), music with prominent staccato passages (e.g., Schlittmeier, Hellbrück, & Klatte, 2008a), and sequences of different tones (e.g., Jones & Macken, 1993) have been shown to impair verbal short-term memory. Conversely, irrelevant sounds with no, or marginal, changing-state features (so-called "steady-state" sounds) do not affect performance, or do so significantly less than do changing-state sounds. For example, music that is played legato reduces performance significantly less than music that is played staccato (Klatte, Kilcher, & Hellbrück, 1995; Schlittmeier et al., 2008a); a repeatedly presented tone impairs performance less than does a sequence of different tones (e.g., Jones & Macken, 1993), and continuous noise does not reduce performance relative to a silent control condition (e.g., Ellermeier & Zimmer, 1997; Jones, Miles, & Page, 1990).

Explanation of the ISE within short-term memory models

The crucial role of a sound's changing-state characteristic for ISE evocation is unquestionable. The changing-state hypothesis proposed by Jones and co-authors has allowed for this role, which is part of the object-oriented episodic record (O-OER) model of short-term memory (e.g., Jones, 1993; Jones, Beaman, & Macken, 1996; Macken, Tremblay, Alford, & Jones, 1999). In addition, two other models

encompass the differing effects of changing-state and steady-state sounds—namely, the feature model (Nairne, 1988, 1990; Neath, 1998, 2000) and primacy model (Page & Norris, 1998, 2003). Please note, the first model to offer a theoretical explanation of the ISE was the working memory model by Baddeley (1986, 2000). Because this model does not explicitly cover the decisive role of a sound's changing-state characteristic or the potential disturbance impact of nonspeech sounds, it will not be considered in the following.

The O-OER model, feature model, and primacy model differ with respect to cognitive mechanisms, which are assumed to underlie the higher disturbance impact of changing-state sounds as compared with steady-state sounds. Within the O-OER model (Jones, 1993; Jones et al., 1996; Macken et al., 1999), an irrelevant changing-state sound results in enhanced interference processes with memory material as compared with a steady-state sound. Here, a changing-state sound is encoded as a sequence of different objects in short-term memory, which are lined up in order of occurrence. This automatically encoded order information interferes with the order information that is stored voluntarily to serially recall the memory items. An irrelevant steady-state sound constitutes the encoding of a single object with only self-referential and, thus, barely concurrent, order information. These assumptions are applied to irrelevant speech and nonspeech sounds within the O-OER model, whereas the feature model restricts interferences as the cause of the ISE to irrelevant speech. In the present study, irrelevant changing-state speech was said to degrade memory representations of to-be-remembered items to a greater extent as compared with steady-state speech. The feature model further assumes a higher burden on attention and processing resources for irrelevant changing-state speech and nonspeech sounds (Neath & Surprenant, 2001), which is, for irrelevant nonspeech, the sole cause of the observed disturbance impact. Finally, the primacy model (Page & Norris, 1998, 2003) unites assumptions that are similar to those of the O-OER model and the feature model by assuming that more order information is encoded for irrelevant changing-state sounds than for steady-state sounds. This results in enhanced consumption of cognitive resources, on the one hand, and more noise among order information encoded for the sequence of memory items, on the other hand (cf. Norris, Baddeley, & Page, 2004).

Finally, each of these three short-term memory models (O-OER model, feature model, and primacy model) refers to a sound's changing-state characteristic to explain its disturbance impact on short-term memory in comparison with less effective steady-state sounds. However, no model to date has provided a criterion to estimate, quantitatively, a given sound's inherent changing-state characteristic. Thus, these short-term memory models have low fidelity in terms

of predicting how much a given sound will disrupt performance.

Considering changing-state as a dimension

To date, changing-state has been described exclusively in qualitative-descriptive terms—namely a changing-state sound, which is a sound that can be perceptually clearly segmented while successive perceptual tokens vary (see the previous section). Consequently, steady-state sounds are considered to lack cues necessary for segmentation (e.g., legato-music, continuous noise) or to consist of a repeated perceptual token (e.g., repeated presentation of a vowel, consonant, or tone). This qualitative description suggests that changing-state and steady-state sounds belong to two disjointed sound categories. Rather, the changing-state characteristic of a sound should be understood in terms of a continuum, with more changing-state features resulting in greater disturbance. In fact, such a dimensional view on changing-state sounds is in accordance with the assumptions of short-term memory models that claim to explain the ISE. Specifically, these models assume only quantitative, but no qualitative, differences, between cognitive mechanisms that are evoked by irrelevant changing-state and steady-state sounds (cf. above).

Several empirical studies have demonstrated different levels of this changing-state. In these studies, the background sound was manipulated in several steps and corresponding gradual variations of its disturbance impact on short-term memory were observed. Specifically, a decrement of the impact of background speech disturbances has been found by enhancing the extent of low-pass filtering (Jones, Alford, Macken, Banbury, & Tremblay, 2000), increasing the level of continuous noise used for superimposition (Ellermeier & Hellbrück, 1998), or by reducing the pitch separation of successive speech tokens (Jones, Alford, Bridges, Tremblay, & Macken, 1999). These manipulations have one thing in common: They reduce the acoustic mismatch between successive items in an irrelevant sound sequence from an auditory-perceptive view and, with this, the prominence of inherent changing-state features.

Speech as maximum disturbance impact

Considering changing-state as a dimension leads to the question of what sounds induce minimum and maximum disturbance impacts on verbal short-term memory. The answer is informative on psychoacoustical measures that might correspond to changing-state characteristics of a given sound and promise to be appropriate for algorithmic modeling of the ISE. Empirical data suggests that the minimum disturbance impact is represented by continuous noise. To our knowledge, no experiment has

verified any significant difference between performance during continuous noise and that during silence (e.g., Ellermeier & Zimmer, 1997; Jones et al., 1990). On the contrary, empirical evidence speaks in favor of speech being the counterpart that induces the maximal disturbance impact.

It has been subject to debate whether nonspeech signals can induce an ISE as high as that induced by background speech (cf., e.g., LeCompte, Neely, & Wilson, 1997; Tremblay, Nicholls, Alford, & Jones, 2000; Jones, Alford, et al., 1999); however, it seems to be unquestioned that narration has never been surpassed in behavioral measures of the ISE. This holds true for our ISE experiments and for all other published studies to date. Since 2001, we have collected behavioral data for 40 background sounds in 70 behavioral measures. Among these sound conditions were different speech signals and nonspeech sounds such as background music, sequences of pure tones, animal sounds, and so on (cf. Table 1). Figure 1 provides a visualization of performance data that was collected in these behavioral experiments with sound conditions sorted in descending order of disturbance impact (See Table 1 for the assignment of sound numbers to sound conditions). Further inspection revealed that background speech elicited the highest level of disturbance. Specifically, 17 of the 20 behavioral measures that were found to have the highest disturbance impact involved irrelevant changing-state speech, including narration (mother tongue or foreign language) or sequences of different consonants. The other three signals were staccato music two times, sounds 8 and 17 in Fig. 1, and one office noise recording, sound 19. Moreover, changing-state speech signals were not found among the other 50 behavioral measures in which performance was reduced to a lesser extent or not reduced at all. Consequently, we assumed speech to be the sound condition that produced the greatest disturbance impact known so far.

The hearing sensation fluctuation strength as a psychoacoustical pendant of changing-state

The sound conditions that induced maximal and minimal ISE (i.e., no disturbance impact) provided information on the psychoacoustical measure that is appropriate for algorithmically modeling the disturbance impact of background sounds. Specifically, continuous noise does not impair performance; its distinguishing mark is a lack of temporal and spectral variations. Conversely, background speech disturbs short-term memory the most; it is characterized by a clear temporal structure constituted by micro- and macropauses and prominent spectral variations between successive auditory–perceptive tokens. The unexcelled disturbance impact of background speech suggests that fluctuation strength is a promising measure for modeling

Table 1 Specification of sound conditions

Sound	Further specifications		Sound Nos. in Fig. 1 with fluctuation strength (*F*) in [vacil] in parentheses
Speech	- Narration	- (A) foreign language or (B) mother tongue	(A) 1 (*1.11*), 2 (*1.21*), 3 (*1.25*), 16 (*0.55*), 33 (*0.26*); (B) 10 (*0.36*), 11 (*1.02*), 12 (*1.13*), 13 (*1.18*), 14 (*1.27*), 15 (*0.87*), 30 (*0.25*)
		- (A) high or (B) low speech intelligibility	(A) 1 (*1.11*), 2 (*1.21*), 3 (*1.25*), 10 (*0.36*), 11 (*1.02*), 12 (*1.13*), 13 (*1.18*), 14 (*1.27*), 11 (*1.02*), 12 (*1.13*), 13 (*1.18*), 14 (*1.27*), 15 (*0.87*), 16 (*0.55*); (B) 30 (*0.25*), 33 (*0.26*)
		- differing levels: (A) 60 dB, (B) 55 dB, or (C) 35 dB	(A) 2 (*1.21*), 3 (*1.25*), 11 (*1.02*), 12 (*1.13*), 13 (*1.18*), 14 (*1.27*); (B) 1 (*1.11*), 15 (*0.87*); (C) 10 (*0.36*), 16 (*0.55*), 30 (*0.25*), 33 (*0.26*)
	- Babble noise (eight speakers)		23 (*0.38*)
	- Sequence of different cv-combinations		4 (*0.94*), 5 (*0.95*), 6 (*0.97*), 7 (*0.95*), 9 (*0.95*), 18 (*0.82*), 20 (*0.80*)
	- Repeated cv-combination or vowel		27 (*0.79*), 36 (*0.80*), 37 (*0.77*), 43 (*0.80*), 47 (*0.80*), 48 (*0.80*), 58 (*0.80*)
Instrumental Music	- Staccato music		8 (*0.49*), 17 (*0.68*), 22 (*0.68*), 32 (*0.68*), 38 (*0.68*)
	- Legato music		41 (*0.20*), 55 (*0.12*), 57 (*0.12*), 63 (*0.12*), 70 (*0.12*)
	- Violin solo		31 (*0.56*)
	- Strings concert		28 (*0.16*), 29 (*0.21*)
Tone Sequences	- Changing-state tones: Sequence of different tones		39 (*0.53*), 40 (*0.53*), 42 (*0.53*)
	- Steady-state tones: Repetition of one tone		21 (*0.32*), 62 (*0.32*), 68 (*0.32*)
Animal Sounds	- Ducks' quacking		34 (*1.77*)
	- Tawny owl		52 (*0.23*)
Traffic Noise	(A) 100 cars/hour at 60 dB, or 2000 cars/hour at (B) 50 dB or (C) 60 dB or (D) 70 dB		(A) 46 (*0.14*); (B) 45 (*0.12*); (C) 60 (*0.10*); (D) 56 (*0.08*)
Office Noise	Recordings of real office noise. Either (A) played-back in isolation or superimposed with (B) staccato-music, (C) legato-music or (D) continuous noise		(A) 19 (*0.41*), 24 (*0.41*), 53 (*0.41*), 54 (*0.46*); (B) 25 (*0.45*), 26 (*0.39*), 44 (*0.39*); (C) 35 (*0.28*), 50 (*0.31*), 59 (*0.28*); (D) 51 (*0.08*), 67 (*0.10*), 69 (*0.08*)
Continuous Noise			49 (*0.03*), 61 (*0.02*), 64 (*0.02*), 65 (*0.02*), 66 (*0.02*)

The database for algorithmic modeling of the ISE encompassed about 70 behavioral measures of short-term memory during about 40 different sounds

the ISE. There are at least two reasons for this notion. First, the hearing sensation fluctuation strength accounts for amplitude and frequency modulations. Second, the perception of fluctuation strength reaches its maximum at 4 Hz. This coincides with a well-known and prominent characteristic of narration: Fluctuations of ~4 Hz in the temporal envelope of speech are due to a syllable frequency of typically about 4 Hz (Fastl, 1987).

Fluctuation strength is perceived when listening to slowly modulated sounds. Up to approximately 20 Hz modulation frequency, the auditory system is able to follow variations in amplitude and frequency of a sound, which results in a fluctuating percept that gives the quantity its name. At modulation frequencies above 20 Hz, the perception of fluctuation strength becomes the hearing sensation roughness. Fluctuation strength reaches a maximum at modulation frequencies of 4 Hz, with the percept

quickly decaying for higher and lower modulation frequencies. From experiments that employed well-defined tone and noise stimuli, it is known that the major determinant factors for fluctuation strength, *F*, are modulation depth (ΔL) and modulation frequency (f_{mod}). This leads to the following simple equation:

$$F \approx \frac{\Delta L}{(f_{mod}/4Hz) + (4Hz/f_{mod})} \quad [\text{vacil}]. \quad (1)$$

The denominator in Eq. 1 assures that *F* reaches a maximum at 4 Hz. The numerator takes into account the modulation depth (ΔL). Note that ΔL does not refer to the physical modulation depth; rather, it refers to the so-called "temporal masking depth" (Fastl & Zwicker, 2010). This temporal masking depth is a measure for perceived modulation depth. Specifically, it takes temporal masking
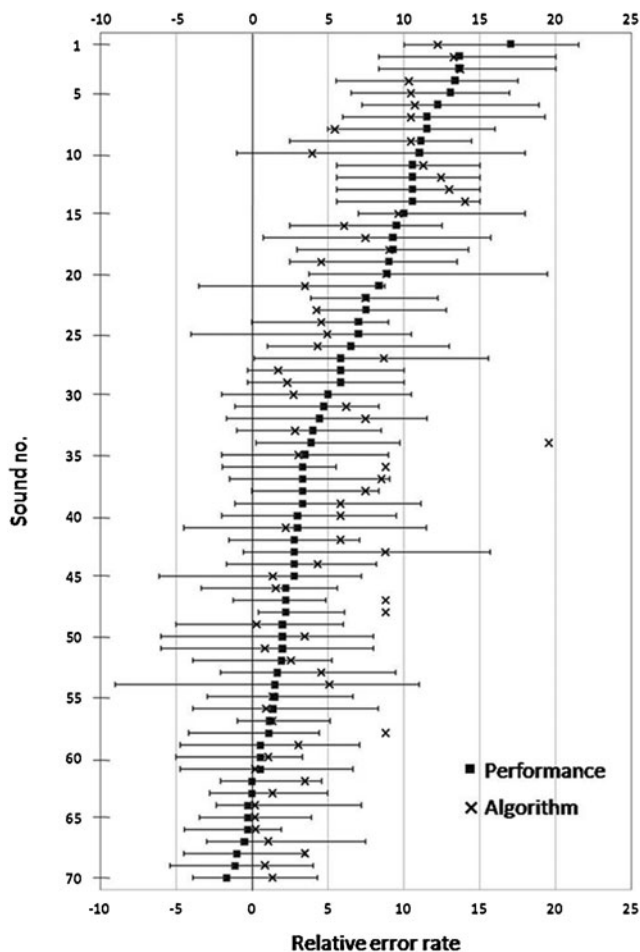
Fig. 1 Verbal short-term memory was explored during different background sounds in 70 experimental measurements. *Squares* represent the relative error rates found in behavioral experiments and are complemented by interquartile ranges. Positive values indicate enhanced error rates during irrelevant sound as compared with performance during silence. *Crosses* represent the algorithmically calculated relative error rates

Research question

The question of interest is whether the hearing sensation fluctuation strength is appropriate to model the ISE induced by speech and nonspeech sounds. If the disturbance impact of a given sound on verbal short-term memory can be linked to this instrumental psychoacoustical measure, then fluctuation strength might be an appropriate measure for the inherent changing-state characteristic of a given background sound. In this case, a criterion for changing-state would be available.

The algorithm

The database for modeling

For the algorithmic modeling of the ISE, a database comprising approximately 40 background sounds and corresponding performance data was used. The experiments were conducted at the Catholic University of Eichstätt-Ingolstadt, beginning in 2001. Of note, some data were collected at the University of Oldenburg in 2001 and 2002. The manipulated background sounds were speech, music, continuous noise, animal sounds, tone sequences, office noise, or traffic noise. Table 1 displays additional information about the sound conditions included in the database. The behavioral data were mainly published in peer-reviewed journals (Schlittmeier & Hellbrück, 2009; Schlittmeier et al., 2008a, 2008b; Schlittmeier, Hellbrück, Thaden, & Vorländer, 2008c) or in full congress contributions (Schlittmeier & Hellbrück, 2004; Schmid & Hellbrück, 2004). Furthermore, only few data were obtained from unpublished diploma or doctoral theses (Hauser, 2005; Schlittmeier, 2005), or have not yet been published.

Participants

The sample sizes for the behavioral experiments ranged from 18 to 36 participants. Participants were students who were enrolled at the Catholic University of Eichstätt–Ingolstadt or the University of Oldenburg. All participants responded to a notice for participation and reported normal hearing. A small allowance or credit points were given for participation.

Procedure

All experiments employed a verbal serial recall task, which is the standard paradigm for measuring verbal short-term memory. The digits 1 to 9 were presented successively in a randomized order. Participants were asked to recall the exact presentation order after a short retention period of

effects, such as post masking into account, which, considering a perceptual perspective, fills the physical dips in a modulated stimulus. Note that, Eq. 1 gives only an estimate. For the final calculation of F, Eq. 1 must be evaluated separately in the different critical frequency bands. These different estimates then must be integrated, considering the boundary condition that a stimulus with 1 kHz and 60 dB (which is fully amplitude modulated at 4 Hz) yields a fluctuation strength of $F = 1$ vacil (vacil denotes the entity of fluctuation strength as defined by Fastl, 1983). In general, F can vary between 0 and approximately 2 vacil, with the latter value being reached only by artificial sounds such as an amplitude-modulated tone. Speech ($F = 1$ vacil) may be the one real-world sound with maximum fluctuation strength. See Table 1 for fluctuation strength values of all sounds in the present database.

10 s. Each trial encompassed an item presentation period, retention period, and recall period. The ISE was defined as the impairment of task performance in the presence of certain irrelevant background sounds as compared with performance in the control condition. The control condition was silence except for one experiment, which used continuous noise as the control. Researchers have demonstrated no impairment in serial recall using this control sound (e.g., Ellermeier & Hellbrück, 1998; Schlittmeier et al., 2008a).

In each experiment included in the database, short-term memory was examined during three to seven different sound conditions. Depending on the experiment, there were 15–20 trials in each sound condition. For each participant, error rates were averaged over all trials in a condition. The participant's performance during silence (baseline) was then subtracted from this value. The resulting value served as a measure of relative error and indicated the participant's performance decrement in the presence of the given background sound. The median of these individual error rates was calculated for each background sound condition.[1] In the following, the median error rates are presented as percentages.

During the test condition, the items to be remembered were presented visually or auditorily. Perfect intelligibility of auditory items in the presence of background sounds was confirmed by speech identification tests. In all experiments, background sounds were presented via loudspeakers or head phones (diotically) at a moderate level of 35–60 dB (A). For loudspeaker presentation, the sound pressure level was measured by a Brüel & Kjær 2231 sound level meter or a NoiseBook 2.0 (HEAD acoustics GmbH). For headphone presentations, a Brüel & Kjær 4153 artificial ear was used.

Estimating a given sound's disturbance impact from its fluctuation strength

For algorithmic modeling of the ISE, the fluctuation strengths of background sounds used in the behavioral experiments included in the database were measured instrumentally. Therefore, the sound files were limited to representative samples of 30 s, which corresponded to the duration of a trial (item presentation, retention interval, and recall). Each sound file was adjusted to the sound pressure level used in the corresponding behavioral ISE experiment.

---

[1] The median is, like the mean, a measure of central tendency. It is the middle score in a set of scores above and below which lie 50% of the other scores. Here, for example, the error rates of all participants during a certain background sound condition are placed in ascending order of size from the smallest to the largest; the median is the middle score. The median is an appropriate measure of central tendency in the present study because the data is not symmetrically distributed in all behavioral experiments.

Then, fluctuation strength for each sound file was measured using the software PAK® (Müller-BBM VibroAkustik Systeme GmbH). This software calculates fluctuation strength according to a model proposed by Aures (1985); a more recent description of this model can be found in Fastl and Zwicker (2010). The model operates on time-varying specific loudness patterns. This model accounts for the fact that, for complex real-world sounds, modulation depth (ΔL cf. the introduction section and Eq. 1) is often not available, but corresponding differences in specific loudness can be used instead. In the present study, calculating the arithmetic mean of fluctuation strength over time led to a single value for each background sound.

A background sound, eliciting error-rates with small interquartile range, was considered an appropriate reference sound condition for the algorithm. Specifically, music with prominent staccato passages (sound 22 in Fig. 1), was chosen for this purpose. The disturbing impact of this background sound was 7.5%, which means that the median of error rates in the verbal short-term memory task increased by 7.5% during staccato music as compared with the silence condition. Short-term memory performance in the presence of the reference sound, in addition to fluctuation strength of the sound (0.68 vacil), was used to normalize the algorithm. The algorithm then estimated the disturbing impact of a given sound on verbal short-term memory (ISE) according to Eq. 2, with F representing the fluctuation strength (in vacil) of the given sound:

$$ISE = \frac{F}{0.68 \; vacil} \cdot 7.5 \quad [\%]. \tag{2}$$

This algorithm was developed at Technische Universität München in a student research paper (first conference presentation: Weißgerber, Schlittmeier, Kerber, Fastl, & Hellbrück, 2008).

Results

Equation 2 was used to calculate the ISE for all 70 experimental measurements. Figure 1 illustrates the estimated and observed error rates for each sound condition.

As indicated in Fig. 1, the algorithm predicted the observed results within the interquartile range in 63 of 70 behavioral measurements. Additionally, the algorithm yielded valid estimates for all sounds that significantly disrupted verbal short-term memory (e.g., speech, staccato music, office noise) as compared with the control condition. The algorithm also adequately modeled the absence of an effect of the continuous steady-state sounds on cognitive performance (e.g., legato music, traffic noise, continuous noise). Despite these findings, the algorithm marginally underestimated the detrimental effect of sound 23, which

was a speech babble noise that was constituted by the superimposition of eight speakers. Furthermore, the algorithm overestimated the effect of six synthetic sounds, which had no significant behavioral effect on short-term memory. These sounds included synthetic steady-state sounds and were characterized by a periodically reoccurring auditory–perceptual event, such as a repeatedly presented vowel (sounds 36, 47, 48, 58), consonant (sound 37), or duck's quacking sounds (sound 34).

Estimated and observed error rates were significantly correlated yielding a Spearman correlation coefficient of $r_s = .74$ ($p < .01$). The corresponding coefficient of determination ($r^2 = .55$) indicated that the estimated and observed error rates shared 55% of the overall variance. These coefficients of correlation and determination were surprisingly high, considering the fact that fluctuation strength was the only auditory–perceptive parameter of a background sound for which the algorithm accounted. Thus, a large part of the variance in performance can be explained by a single psychoacoustic measure.

Discussion

The ISE describes the empirically robust finding that background sounds, with distinct temporal–spectral variations, significantly reduce verbal short-term memory. Although researchers have widely explored this phenomenon in behavioral experiments, predicting the effect of a given sound on cognitive performance, to date, has not been possible. As such, the present algorithm provides the missing link between an instrumentally measurable psychoacoustical parameter of an irrelevant sound and its impact on verbal short-term memory.

Fluctuation strength indicates the ISE caused by irrelevant speech and nonspeech

Based on the hearing sensation fluctuation strength, the current algorithm correctly modeled 63 of 70 behavioral ISE measurements within the interquartile ranges. Specifically, the performance impact of all background sounds in the database, which are found in real-world settings, were estimated correctly. These findings apply to performance-reducing changing-state sounds, which include background speech, staccato music, and office noise, as well as nondisturbing continuous steady-state sounds, such as legato music and traffic noise. Six of the seven sounds that the current algorithm did not adequately estimate included synthetic sounds, characterized by a single, periodically reoccurring auditory–perceptual token (e.g., a consonant or a vowel). Such steady-state sounds have marginal or no effect on short-term memory in behavioral experiments (see the introduction section); however, the algorithm over-

estimated the impact of these sounds. This overestimation occurred because the algorithm does not "hear" that successive auditory-perceptual tokens are identical (e.g., A A A A A A) or different (e.g., A I U O E I) from each other. Rather, such repeating and nonrepeating sequences differ in fluctuation strength only marginally. In the future, a refinement of the present algorithm may be accomplished so that such artificial steady-state sounds are estimated correctly. This might be realized by introducing a periodicity or pattern identifier to correct for the steady-state that is constituted by the repetition of an auditory–perceptive token.

Nonetheless, already in its present form, the algorithm satisfactorily predicted the detrimental impact of background speech and nonspeech sounds on short-term memory. With this, the hearing sensation fluctuation strength outperformed the first ambitious efforts to relate instrumentally measurable sound characteristics and cognitive performance by Hongisto (2005). The author aimed to describe the quality of acoustics in open-plan offices via the speech transmission index (STI), an instrumental measure of speech intelligibility in rooms (Steeneken & Houtgast, 1980). Though, the STI is an index of speech intelligibility and cannot account for the effects of irrelevant nonspeech sounds (e.g., nonspeech office noise, background music, etc.). Because office employees report background speech as the most disturbing background sound, this restriction might not be a crucial point to evaluate office acoustics (cf. Haka et al., 2009). However, even for background speech, applying fluctuation strength might be of greater precision than applying the STI in certain cases. For example, the STI cannot adequately estimate the intelligibility of extremely low-pass filtered and low-level speech signals as tested by Schlittmeier et al. (2008c). In this study, the level of speech was reduced to 35 dB(A) either in a broadband filter condition (speech intelligibility remained high) or in a low-pass filter condition with 60 dB(A) attenuation around 4,000 Hz (low speech intelligibility), and verbal short-term memory was measured in these background speech conditions. For the intelligibility of such speech signals, however, various factors, such as decreased modulation depth and different modulation perceptions near the listening threshold, play a role that is not covered by the STI. In contrast, the present algorithm accurately estimated the succession of sound conditions with respect to induced error rates. In more detail, the algorithm reproduced that unaltered speech induced the highest error rates, followed by speech that was reduced in level. A combined reduction of speech level and intelligibility reduced error rates; however, the algorithm expected still more errors than during the silent control condition as was observed in the performance experiment (cf. Schlittmeier et al., 2008c).

Does fluctuation strength indicate a sound's inherent changing-state characteristic?

The present investigation revealed that a background sound's fluctuation strength and disturbance impact are highly correlated; specifically, with rising fluctuation strength, disturbance impact increased. Consequently, this hearing sensation is suggested as an external criterion that allows predictions on the degree a given background sound disturbs short-term memory. Nonetheless, the algorithm might not suffice in providing a psychoacoustical measure for a sound's changing-state characteristic.

As was described in the introduction section, earlier studies have developed the changing-state definition to include two criteria: (a) segmentability from a temporal–spectral perspective, and (b) successive auditory–perceptive tokens being different from each other. Fluctuation strength appears to account for the first changing-state criterion because of its increment with rising modulation depth, ΔL, in the temporal envelope (cf. the introduction section, Eq. 1). However, the second changing-state criterion cannot be evaluated by this hearing sensation because the considered parameters (i.e., modulation depth, ΔL, and modulation frequency, f_{mod}) do not allow for identifying whether successive auditory–perceptive tokens are different. For the same reason, the algorithm could be expected to emulate incorrect behavioral data that demonstrates the role of auditory–perceptive grouping processes in ISE evocation. Such experiments have shown that the two changing-state criteria (temporal segmentation, different tokens) must be fulfilled within one coherent auditory-perceptual stream to evoke an ISE. Imagine, for example, a high and a low tone that alternates by a frequency of 2 Hz (i.e., two tones per second). For small to moderate frequency differences between the two tones, one sequence of two alternating tones is perceived (high–low–high–low–high, etc.). If the frequency deviance between the two tones exceeds about 10 semitones, the percept of 1 two-token sequence turns into the percept of 2 one-token sequences (high–high–high, etc.; low–low–low, etc.; van Noorden, 1975). The latter two steady-state sequences, however, reduced short-term memory significantly less than the changing-state sequence constituted by two alternating tones (Jones, Alford, et al., 1999). According to the hearing sensation fluctuation strength, the present algorithm does not account for these differential sound effects due to primitive auditory streaming by pitch (cf. Bregman, 1990; further example: streaming by location; Jones, Saint-Aubin, & Tremblay, 1999).

Even if the algorithm simply represents an index of segmentability, it succeeds for 90% of the sounds in the database and, in particular, on all real-world sounds. Thus, the algorithm might serve in the future as a component of a more complete model because it models one necessary sound characteristic (segmentability) for ISE evocation, even though it must be complemented by modeling of further characteristics (i.e., successive segments must differ from each other, auditory–perceptual streaming processes) to account for all aspects of the phenomenon. In addition, experimental validation of the potential cause–effect relationship between fluctuation strength and disturbance impact is needed. This might be accomplished by systematically varying the fluctuation strength of a given background sound in several steps while measuring the impact of these sounds on performance. A procedure used by Ellermeier and Hellbrück (1998) might be appropriate for this purpose. Specifically, the authors superimposed irrelevant speech by continuous noise of different signal-to-noise ratios (SNR). With decreasing SNR (i.e., noise becomes louder than the speech signal) memory impairment decreased. This result is most probably due to the fact that continuous noise partially masks (i.e., makes less perceptible) soft, yet necessary, cues for the segmentation of speech (e.g., consonants) and fills up micro- and macro-pauses in the speech signal. This is accomplished more efficiently as louder noise is compared to the speech signal (i.e., the smaller the SNR). With decreasing SNR, however, fluctuation strength decreases and corresponding performance effects can be explored. In a replication of the aforementioned experiment (Ellermeier & Hellbrück, 1998), which includes an extension to nonspeech sounds (e.g., music, office noise, or sequences of different tones), behavioral testing must show whether the disturbance impact of these signals on verbal short-term memory vary with their fluctuation strength.

A special role for speech in ISE evocation?

In the present database, background speech induced the highest disturbance impact on behavioral measures of verbal short-term memory. Undoubtedly, speech plays a significant role in human cognition and information processing. Furthermore, it is not farfetched to assume that speech plays a special role with respect to its detrimental impact if heard in the background while performing a cognitive activity. However, whether speech-like qualities of an irrelevant sound are relevant with respect to ISE evocation is still debatable (cf., e.g., Jones et al., 1992; Jones & Macken, 1993; LeCompte et al., 1997; Salamé & Baddeley, 1989). Whereas the O-OER model (e.g., Jones & Macken, 1993) assumes irrelevant speech and nonspeech to function equivalently, the feature model (e.g., Neath & Surprenant, 2001) and the primacy model (e.g., Page & Norris, 1998, 2003) ascribe speech as having a qualitatively different role than nonspeech signals in ISE evocation (cf. the Introduction section).

One might assume that two results of the present study are informative concerning a special role for speech. Firstly, the algorithm verified that fluctuation strength is suitable for modeling the detrimental impact of both speech and nonspeech signals. However, the signals were not matched for fluctuation strength; rather, the nonspeech signals included in the database yielded lower fluctuation strengths than did the changing-state speech signals. Thus, it cannot be inferred whether irrelevant speech and nonspeech function equivalently (O-OER model) or qualitatively different (feature model, primacy model). Secondly, speech was the background sound with a maximum disturbance impact. This result is consistent with both the feature model and the primacy model; however, it could also fit into the O-OER model. This is the case when speech is the signal with the highest changing-state tested because the O-OER model attributes a sound's disturbance impact exclusively to its inherent changing-state characteristic.

The question arises, why the behavioral effect of irrelevant speech is unexcelled. In our opinion, this is most likely because speech signals best fulfill the criteria for sounds to gain automatic and obligatory access to short-term memory. At first, heard speech is just a complex time-varying acoustic signal. The perceived sequences of auditory–perceptive tokens may convey meaning; however, because of the sequential nature of speech, higher level interpretations often necessitate temporary and sequential storage of auditory input (cf. Davis & Johnsrude, 2007). Thus, it might be reasonable to assume that humans are not able to prevent heard speech from gaining access to short-term memory; however, its semantic content is not a suitable access criterion. Such criteria might include clear temporal structures and significant changes in spectrum as the applicability of the hearing sensation fluctuation strength for modeling the ISE suggests.

Heard speech, no doubt, is distinguished by more than just spectral and temporal variations. Obviously, a speech signal's semantic content is not accounted for by the instrumental measure of fluctuation strength. Several ISE studies have found a contribution of semantics to error rates (Buchner & Erdfelder, 2005; Buchner, Rothermund, Wentura, & Mehl, 2004; LeCompte et al., 1997; but cf. Jones et al., 1990, and Klatte et al., 1995, for different results). This finding, however, is most likely due to attention capture adding to short-term memory interference (i.e., the ISE; cf. Klatte, Lachmann, Schlittmeier, & Hellbrück, 2010, for a discussion). Thus, a semantic effect is not in contradiction to the proposal that the fluctuation strength of an irrelevant sound is the crucial factor for ISE evocation. Apart from the question of what the underlying cognitive mechanisms of this semantic effect are, semantics, obviously, cannot contribute to the disturbance impact of many background sounds that have been found to impair short-term memory due to these sounds' lacking any semantic content—that is, sequences of different consonants (e.g., Jones et al., 1992), foreign language (e.g. Colle &

Welsh, 1976), and nonspeech sounds (e.g. Schlittmeier et al., 2008a). Conversely, fluctuation strength is an overarching sound parameter that can be instrumentally calculated for all potential background sounds and does not apply to only a subgroup of sounds. And it does account for a large part of observed performance variance during real-world background sounds, whether speech or nonspeech. The important strengths of the proposed algorithm lie, in particular, in these aspects.

References

Aures, W. (1985). Ein Berechnungsverfahren der Rauhigkeit [A procedure for calculating roughness]. *Acustica, 58,* 268–281.

Baddeley, A. (1986). *Working memory.* Oxford, England: Clarendon.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Bregman, A. S. (1990). *Auditory scene analysis. The perceptual organization of sound.* Cambridge, MA: MIT Press.

Buchner, A., & Erdfelder, E. (2005). Word frequency of irrelevant speech distractors affects serial recall. *Memory & Cognition, 33,* 86–97.

Buchner, A., Irmen, L., & Erdfelder, E. (1996). On the irrelevance of semantic information for the "Irrelevant Speech" effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49,* 765–779.

Buchner, A., Rothermund, K., Wentura, D., & Mehl, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition, 32,* 722–731.

Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior, 15,* 17–31.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influence on the interface of audition and speech perception. *Hearing Research, 229,* 132–147.

Ellermeier, W., & Hellbrück, J. (1998). Is level irrelevant in 'Irrelevant Speech'? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology. Human Perception and Performance, 24,* 1406–1414.

Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the "irrelevant speech effect". *Journal of the Acoustical Society of America, 102,* 2191–2199.

Fastl, H. (1983). Fluctuation strength of FM-tones. In *Proceedings of the 11th International Congress of Acoustics* (pp. 123–126). Paris, France: ICA.

Fastl, H. (1987). A background noise for speech audiometry. *Audiological Acoustics, 26,* 2–13.

Fastl, H., & Zwicker, E. (2010). *Psychoacoustics—facts and models* (3rd ed.). Heidelberg, Germany: Springer.

Haka, M., Haapakangas, A., Keränen, J., Hakala, J., Keskinen, E., & Hongisto, V. (2009). Performance effects and subjective disturbance of speech in acoustically different office types—a laboratory experiment. *Indoor Air, 19,* 454–467.

Hauser, C. (2005). *Lernen bei sprachlichem und nichtsprachlichem Hintergrundschall - der Einfluss des zeitlichen Schallverlaufs auf die Behaltensleistung im Kurzzeitgedächtnis.* Unpublished diploma thesis. Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany.

Hellbrück, J., & Liebl, A. (2008). Noise effects on cognitive performance. In S. Kuwano (Ed.), *Recent topics in environmental psychoacoustics* (pp. 153–184). Osaka, Japan: Osaka University Press.

Hongisto, V. (2005). A model predicting the effect of speech of varying intelligbility on work performance. *Indoor Air, 15,* 458–468.

Jones, D. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 87–104). Oxford, England: Clarendon.

Jones, D. M., Alford, D., Bridges, A., Tremblay, S., & Macken, B. (1999a). Organizational factors in selective attention: The interplay of acoustic distinctiveness and auditory streaming in the irrelevant sound effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 464–473.

Jones, D. M., Alford, D., Macken, W. J., Banbury, S., & Tremblay, S. (2000). Auditory distraction from degraded stimuli: Linear effects of changing-state in the irrelevant sequence. *Journal of the Acoustical Society of America, 108,* 1082–1088.

Jones, D. M., Beaman, P., & Macken, W. J. (1996). The object–oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short–term memory* (pp. 209–237). Hove, England: Psychology Press.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 369–381.

Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44,* 645–669.

Jones, D. M., Miles, C., & Page, J. (1990). Disruption of proofreading by irrelevant speech: Effects of attention, arousal or memory? *Applied Cognitive Psychology, 4,* 89–108.

Jones, D. M., Saint-Aubin, J., & Tremblay, S. (1999b). Modulation of the irrelevant sound effect by organizational factors: Further evidence from streaming by location. *Quarterly Journal of Experimental Psychology, 52A,* 545–554.

Klatte, M., Kilcher, H., & Hellbrück, J. (1995). Wirkungen der zeitlichen Struktur von Hintergrundschall auf das Arbeitsgedächtnis und ihre theoretischen und praktischen Implikationen. *Zeitschrift für Experimentelle Psychologie, 42,* 517–544.

Klatte, M., Lachmann, T., Schlittmeier, S., & Hellbrück, J. (2010). The irrelevant sound effect in short-term memory: Is there developmental change? *European Journal of Cognitive Psychology, 22,* 1168–1191.

LeCompte, D. C., Neely, C. B., & Wilson, J. R. (1997). Irrelevant speech and irrelevant tones: The relative importance of speech to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 472–483.

Macken, W., Tremblay, S., Alford, D., & Jones, D. (1999). Attentional selectivity in short-term memory: Similarity of process, not similarity of content, determines disruption. *International Journal of Psychology, 34,* 322–327.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16,* 343–352.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Neath, I. (1998). *Human memory. An introduction to research, data, and theory.* Pacific Grove, California: Brooks/Cole.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7,* 403–423.

Neath, I., & Surprenant, A. M. (2001). The irrelevant sound effect is not always the same as the irrelevant speech effect. In H. L. I. Roediger & J. S. Nairne (Eds.), *Nature of remembering: Essays in honor of Robert G. Crowder* (pp. 247–265). London, England: American Psychological Association.

Norris, D., Baddeley, A. D., & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1093–1105.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Page, M. P. A., & Norris, D. G. (2003). The irrelevant sound effect: What needs modelling, and a tentative model. *Quarterly Journal of Experimental Psychology, 56A,* 1289–1300.

Salamé, P., & Baddeley, A. D. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24,* 263–265.

Salamé, P., & Baddeley, A. D. (1987). Noise, unattended speech and short-term memory. *Ergonomics, 30,* 1185–1194.

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41,* 107–122.

Schlittmeier, S. (2005). *Arbeitsgedächtnis und Hindergrundschall. Gibt es einen Irrelevant Sound Effect bei auditiv präsentierten Items?* Berlin, Germany: Logos.

Schlittmeier, S. J., & Hellbrück, J. (2004). Noise impact on mental performance. Implications for sound environments at workplaces. *18th International Congress on Acoustics (ICA 2004)* in Kyoto, Japan. Paris, France: ICA.

Schlittmeier, S. J., & Hellbrück, J. (2009). Background music as noise abatement in open-plan offices: A laboratory study on performance effects and subjective preferences. *Applied Cognitive Psychology.* doi:10.1002/acp.1498

Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008a). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology, 20,* 252–271.

Schlittmeier, S. J., Hellbrück, J., & Klatte, M. (2008b). Can the irrelevant speech effect turn into a stimulus suffix effect? *Quarterly Journal of Experimental Psychology, 61,* 665–673.

Schlittmeier, S. J., Hellbrück, J., Thaden, R., & Vorländer, M. (2008c). The impact of background speech varying in intelligibility: Effects on cognitive performance and perceived disturbance. *Ergonomics, 51,* 719–736.

Schmid, A., & Hellbrück, J. (2004). Detrimental effects of traffic noise on basic cognitive performance. In *Proceedings of the Joint Congress CFA/DAGA 04* [CD-ROM]. Oldenburg, Germany: DEGA.

Steeneken, H. J., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *Journal of the Acoustical Society of America, 67,* 318–326.

Tremblay, S., Nicholls, A. P., Alford, D., & Jones, D. M. (2000). The irrelevant sound effect: Does speech play a special role? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1750–1754.

van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences.* Technische Hogeschool Eindhoven, Eindhoven, Diss.

Weißgerber, T., Schlittmeier, S., Kerber, S., Fastl, H., & Hellbrück, J. (2008). Ein Algorithmus zur Vorhersage des Irrelevant Sound Effects. In U. Jekosch and R. Hoffmann (Eds.), *Tagungsband Fortschritte der Akustik—DAGA 2008, Dresden* (pp. 185–186). DEGA, 2008. 10.-13.03.2008. Oldenburg, Germany: DEGA.