Part VI
Free Talk Session 3

REPRESENTATIONAL MOMENTUM AND ANISOTROPY IN VISUAL SPACE

Timothy L. Hubbard[1] and Susan E. Ruppel[2]
[1] *Arizona State University, Tempe AZ 85281 USA*
[2] *University of South Carolina Upstate, Spartanburg, SC 29303 USA*
<timothyleehubbard@gmail.com>

The possibility of anisotropies in visual space in and near the final location of a moving target was examined. Experiments 1 and 2 presented a moving target, and after the target vanished, participants indicated the final location of the leading or trailing edge of the target. Memory for both edges was displaced forward from the actual final locations, and the magnitude of displacement was smaller for the leading edge. Experiments 3 and 4 also presented stationary objects in front of and behind the final location of the target, and participants indicated the location of the nearest or farthest edge of one of the stationary objects. Memory for the near or far edge of an object in front of the target was displaced backward, and memory for the near or far edge of an object behind the target was displaced forward; the magnitude of displacement was larger for objects in front of the target and when the edge was farther away. The findings (a) suggest representational momentum is associated with an anisotropy of visual space that extends across and outward from the moving target and (b) are consistent with previous findings regarding estimation of time-to-contact, anorthoscopic perception, and memory psychophysics.

EBBINGHAUS ILLUSION IN CONTRAST-DEFINED AND TEXTURE-DEFINED STIMULI

Sofia Lavrenteva and Ikuya Murakami
Department of Psychology, the University of Tokyo, Tokyo, Japan

Shape perception in our vision originates not only from edges defined by the first-order information, i.e., luminance difference, but also from edges defined by the second-order information, such as luminance contrast and texture differences. Here we ask whether shape from these types of information is processed by a common mechanism. To address this issue, we investigated whether a geometrical-optical illusion occurs under conditions of isoluminance when only second-order information is available. We used Ebbinghaus illusion figures that were comprised either of sine-wave gratings or uniform gray disks. The luminance, contrast and grating orientation of the central target were manipulated in order to create stimuli defined by luminance, contrast or texture. The surrounding inducers were always disks filled with uniform gray. We found that the illusion occurred robustly under all of these conditions, which supports the notion of the shared mechanism for shape perception in the first-order and the second-order systems.

INTELLIGIBILITY OF ENGLISH NOISE-VOCODED SPEECH RESYNTHESIZED FROM SPECTRAL-CHANGE FACTORS

Takuya Kishida
Graduate School of Design, Kyushu University, Fukuoka, Japan

Yoshitaka Nakajima, Kazuo Ueda, and Gerard B. Remijn
Department of Human Science/Research Center for Applied Perceptual Science
Kyushu University, Fukuoka, Japan

Sophia Arndt and Mark A. Elliott
School of Psychology, National University of Ireland Galway, Galway
Republic of Ireland

Previous studies had found that intelligible Japanese speech sounds can be resynthesized from 3 factors extracted with factor analysis of spectral changes of speech sounds [Kishida et al. (2016). Front. Psychol., 7, 517]. These factors consistently appear among different languages [Ueda and Nakajima. (2017). Sci. Rep., 7, 42468]. Here, we investigated whether and how well these spectral-change factors convey linguistic information of British English. We first performed factor analyses of British-English speech. Nine sets of spectral-change factors having different numbers of factors were extracted. Following this, we resynthesized the speech sounds from the extracted factors as noise-vocoded speech. Five men and four women, ranging in age from 20 to 42 years (mean age = 24.4 years, SD = 6.5 years), participated in the intelligibility test of the resynthesized speech. They were all native speakers of Irish English. The participant wrote down what he/she could hear after listening to three repetitions of the speech stimuli. The word identification scores were considerably lower than the mora identification scores in the previous intelligibility test of Japanese noise-vocoded speech. The present performance might have been affected by a higher speech rate. Consistent with the previous test, 3 factors or more were necessary to make the speech stimuli understandable: When the number of factors was 3, the word identification was 23%. The identification performances gradually improved up to 81.6% when the number of increased up to 9. The results provide support, although limited, for the importance of 3 or 4 factors in British English. [This study was supported by the JSPS KAKENHI Grant Number 16J05172.]

IRRELEVANT SOUND EFFECTS WITH LOCALLY TIME-REVERSED SPEECH

Wolfgang Ellermeier and Florian Kattner
Technische Universität Darmstadt, Germany

Kazuo Ueda and Yoshitaka Nakajima
Kyushu University, Japan

<[ellermeier, kattner]@psychologie.tu-darmstadt.de
[ueda, nakajima]@design.kyushu-u.ac.jp>

Abstract

Irrelevant sound experiments were performed using three versions of temporally altered speech [Ueda et al. Nature Scientific Reports (2017)]: (1) locally time-reversed speech for which recordings were inverted in segments of 20, 70, or 120 ms duration, (2) reverse playback of thus generated material, and (3) the original recording played backwards. N=38 participants were exposed to these sounds derived from German free-running speech, and another N=43 to the same varieties derived from Japanese—supplemented by control conditions with unaltered speech or noise—while they performed a serial recall task. All were native speakers of German, but did not understand Japanese. The results show significant effects of irrelevant speech on memory which are exacerbated when the utterances become intelligible, namely with the original (German) recording or locally time-reversed speech of short segment duration (20 ms). These effects are attenuated when the distractor is derived from Japanese, unintelligible to our participants. The results are interpreted in terms of phonetic and prosodic processing of the irrelevant sound.

Overhearing irrelevant background speech is known to interfere with the maintenance of information in short-term memory, i.e. the 'irrelevant speech effect' (Banbury, Macken, Tremblay & Jones, 2001; Ellermeier & Zimmer, 2014). To demonstrate that the semantics of the irrelevant stream are not crucial to obtain it, some researchers have used 'reversed speech' (Jones, Miles & Page, 1990; LeCompte, Neely & Wilson, 1997) by playing long utterances from end to beginning, thereby maintaining the overall spectrum and the kinds of spectral changes occurring, but degrading the temporal waveform. Typically, irrelevant speech effects of the same magnitude as with 'forward' speech were obtained; see, however Viswanathan, Dorsi, & George (2014).

Rather than playing an entire speech utterance backwards, the present research focuses on 'locally time-reversed' speech, i.e. temporally inverting short successive segments (20–170 ms long) of an extended speech utterance. Research from our laboratories (Ueda, Nakajima, Ellermeier & Kattner, 2017) assessing the intelligibility of thus locally time-reversed speech had shown it to decrease from perfectly intelligible (at 20 ms segment duration) to practically incomprehensible (> 100 ms), in a very similar fashion for four different languages studied (English, German, Mandarin Chinese, and Japanese).

To investigate, how altering temporal detail in the interfering auditory stream might affect the magnitude of the irrelevant speech effect, we thus used (a) the original recordings, (b) locally time-reversed speech, (c) backward speech, and (d) reversed playback of the entire locally time-reversed sequence—thereby concatenating recovered 'forward' segments in an unnatural order—as distractors in the irrelevant speech paradigm (see Fig. 1 for depictions of these stimulus manipulations). We conjectured that (Hypothesis 1) by degrading 'local phonetics' through time reversal of short segments of the signal, (b) should produce smaller irrelevant speech effects than (d) where the integrity of the speech elements is reinstated (depending, of course, on the duration of the reversed segments). Furthermore, if the large-scale pitch changes in the course of an utterance played a role, we expected that (Hypothesis 2) conditions in which this 'global prosody' is preserved (a and b), would produce greater irrelevant speech effects than conditions in which it is destroyed (c and d). Finally, we wanted to study the effects of our time-reversal manipulations depending on whether the irrelevant stream is derived from the listener's native language, or not.

Method

Participants

Two samples of participants were recruited from the same population: $N = 38$ (10 male, age range 17-43 years, MD = 20) were exposed to German irrelevant speech, $N = 43$ (24 male, age range 18-56 years, MD = 23) to Japanese. The majority consisted of university students participating for course credit; the remainder was paid a honorarium of 8 Euros. All of the participants claimed to have normal hearing.

Apparatus and Stimuli

The speech material (German or Japanese) was extracted from a multilingual database of spoken sentences (NTT-AT, 2002) recorded with a 16-kHz sampling rate and 16-bit quantization. For each trial, some 6-8 sentences were concatenated to produce the desired 14-s irrelevant-speech streams. These were processed as illustrated in Fig. 1: They were either played back as such (original forward speech) or divided up into segments of 20, 70, or 120 ms duration, including 2.5-ms cosine ramps to fade in and out. Subsequently, each segment was reversed in time while maintaining the original order of segments, thus generating 'locally time-reversed' speech streams (depiction (b) in Fig. 1). Additional background speech conditions were produced by 'globally reversing' the thus generated material, i.e. having 'reversed speech' (when the original is played backwards in its entirety), or 'reversed' playback of the previously segmented and locally inverted material ('Rev[ltr]'; see (d) in Fig. 1). For each of the resulting eight 'irrelevant-speech' conditions, 10 different exemplars were generated, thus presenting new acoustical material on each trial.

The stimuli were D/A converted by a high-quality sound card (RME multiface II), subsequently passed through a headphone amplifier (Behringer Pro 8) and diotically delivered to electrodynamic headphones (Beyerdynamic DT 990). Sound levels were adjusted using a 74-dB SPL, 1-kHz calibration tone (supplied with the test sentence database) the level of which was measured at the headphones using an artificial ear (Brüel & Kjær type 4153) fitted with a condenser microphone (Brüel & Kjær type 4192), and connected to a sound level meter (Brüel & Kjær type 2250). The experiment was conducted in a double-walled, sound-attenuated chamber (Industrial Acoustics Company).

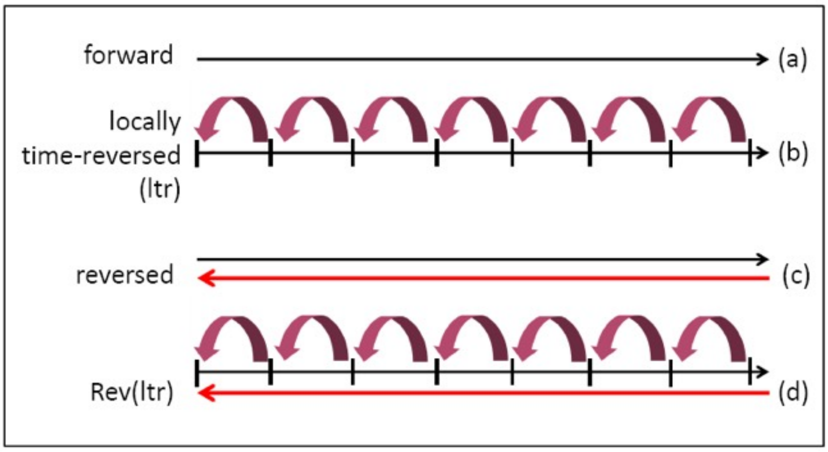

Fig. 1. Schematic representation of the temporal signal processing applied. Top two drawings: Speech utterances were either played forward (as recorded), or locally reversed in time ('ltr'; with segment durations of 20, 70, and 170 ms; these differences not being depicted here). Bottom two drawings: When the entire signal thus obtained was played from end to beginning, reversed speech, or Rev(ltr) conditions resulted.

Procedure

Each trial was initiated by the participant pressing a button. Then, a random permutation of eight digits (drawn from the set of 1 through 9 without repetition) was presented in the center of the screen. Digits were displayed for 1 s each without inter-stimulus intervals. The participants' task was to memorize the order of these digits. After presentation of the digits, a blank screen appeared for 6 s (retention interval) before participants were asked to recall the series of digits. To that effect, a number pad was presented on the screen, and participants had to click on the digits in the respective order. The stream of irrelevant sound (14 s) was played back during the presentation of the digits and during the retention interval. Each background condition (8 derived from speech utterances plus pink noise as a control) was presented 10 times, resulting in a total of 90 trials. The order of trials was randomized for each participant. The experiment proper was preceded by two practice trials, and interrupted by three optional breaks. It took about 60 min to complete.

Results

Performance, i.e. proportion of digits correctly reported at the appropriate position in the sequence, in each of the 9 irrelevant-sound conditions was averaged within and across subjects and is depicted in Fig. 2. The results with German irrelevant speech are depicted in Fig. 2A, the results using Japanese irrelevant speech in Fig. 2B. It is evident that with both background languages, substantial irrelevant speech effects are obtained: Performance in all conditions involving speech (processed or not) is considerably worse than in the pink-noise control conditions (upward-pointing triangles in the two graphs); all $p < .01$ in pairwise Bonferroni-corrected t-tests. Overhearing unprocessed (forward) speech (right-most open circles) appears to constitute the floor for serial-recall performance, and the remaining effects (using temporally processed speech) vary within a range approximately

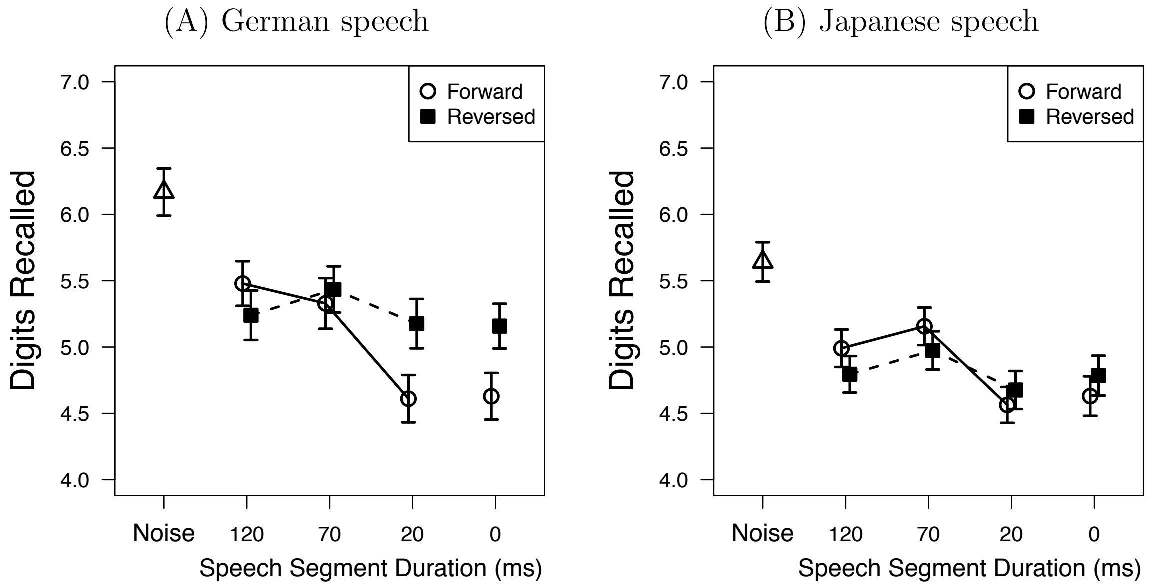

Fig. 2. Performance under irrelevant speech conditions including a 'pink-noise' control condition, 'forward' speech (rightmost open circles in the two graphs), or various conditions derived from locally time-reversed speech (the connected data points), either played as generated (forward) or from end to beginning (reversed). Playing the entire utterance backwards is symbolized by the solitary filled squares at the right of each graph. Means are depicted along with standard errors. The left graph (A) refers to irrelevant sound derived from German speech, the right graph (B) to sound derived from Japanese and presented to $N = 38$ or $N = 43$ German participants, respectively.

covering one-half of the total irrelevant speech effect.

Of primary interest, of course, are the effects of varying segment length and the direction of playback (i.e. ltr and forward speech vs. R(ltr) and globally reversed speech). Therefore, two separate 2 (DIRECTION) $\times$ 4 (SEGMENT LENGTH) repeated-measures ANOVAs were performed on the Japanese and German-language data sets, involving all irrelevant speech conditions except for the pink-noise controls. For the German speech material (Fig. 2A), both a significant main effect of SEGMENT LENGTH, $F(3, 111) = 15.11$; $p < .001$, and of reversing DIRECTION emerged, $F(1, 37) = 11.02$; $p < .01$. More importantly, the interaction between the two factors was significant, $F(3, 111) = 8.54$; $p < .001$, suggesting that the effect of segment duration depends on the direction of playback (or that the solid and dashed curves in Fig. 2A are statistically distinct). Post-hoc testing reveals that for the 20-ms segment duration, and for the 'original' (0 ms segment duration), the direction of playback matters ($p = .044$ and $p = .011$, respectively).

The pattern is different when the Japanese-language irrelevant speech effects (observed with German listeners) are analyzed (Fig. 2B): Here, the main effect of SEGMENT LENGTH is statistically significant, $F(3, 126) = 9.88$; $p < .001$, while the main effect of DIRECTION is not $F(1, 42) = 0.21$; $p = .096$; nor is their interaction, $F(3, 126) = 2.16$; $p = .096$, suggesting that the dashed and solid curves in Fig. 2B are statistically indistinguishable. Notably, playing the entire recording forwards or backwards (data points at 0 ms segment duration) makes no difference for the magnitude of the irrelevant speech effect ($p > .99$).

Discussion

The present study shows that sizeable irrelevant speech effects may be obtained with locally time-reversed speech. These effects reach their maximum when the segment duration is short (20 ms), i.e. when the utterances become intelligible (Ueda, Nakajima, Ellermeier & Kattner, 2017). The effect of segment duration, however, is more pronounced in a language the participants understand (German, in the present case). With a foreign language in the irrelevant stream (here: Japanese), the effect of segment duration appears attenuated, as if listeners cannot make sense of the phonetic elements heard in the background. Therefore, successively degrading the local phonetics of the speech stream (by temporally reversing longer and longer signal segments) appears to reduce the irrelevant speech effect to approximately half its original size, and particularly so in a language the listener is accustomed to process. Our initial hypothesis 1, however, that reinstating the integrity of local speech elements, as in 'Rev(ltr)', would produce greater irrelevant speech effects (than 'ltr'), does not appear to be tenable: Rather, at short segment duration, the opposite occurs.

As for hypothesis 2: Reversing the entire sequence (i.e. playing back from end to beginning, see the two bottom depictions in Fig. 1) which was thought to affect the (global) prosody of the utterance, was indeed effective, though only with the familiar language (Fig. 2A). Here, altering global prosody appears to render the percept more 'noise-like', and that is true for the shortest segment duration [Rev(ltr20)] as well as for the entire stream (0 ms segment duration) played backwards.

Clearly, one would like to see further confirmation of these effects by reversing the roles of the two languages studied, i.e. by presenting the same material to Japanese listeners. That is done in a companion paper to appear in the same proceedings.

Acknowledgements

This research constitutes collaborative work jointly performed at TU Darmstadt and Kyushu University. It was partly sponsored by the Japan Society for the Promotion of Science (JSPS). The authors would like to thank Karla Salazar Espino for collecting the German-language data as part of her bachelor thesis, and Maria Hernando for running the Japanese-language conditions as a graduate research assistant.

References

Banbury, S. P., Macken, W. J., Tremblay, S., & Jones, D. M. (2001). Auditory Distraction and Short-Term Memory: Phenomena and Practical Implications. *Human Factors*, 43(1), 12–29. DOI:10.1518/001872001775992462

Ellermeier, W. & Zimmer, K. (2014). The psychoacoustics of the irrelevant sound effect. *Acoustical Science and Technology*, 35(1), 10–16. DOI:10.1250/ast.35.10

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.

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, Mempory, and Cognition*, 23, 1–12.

NTT-AT (2002). Multilingual speech database 2002 (NTT Advanced Technology Corporation, Tokyo, Japan). URL: http://www.ntt-at.com/product/speech2002/ (date last viewed 6/08/17)

Ueda, K., Nakajima, Y., Ellermeier, W., & Kattner, F. (2017). Intelligibility of locally time-reversed speech: A multilingual comparison. *Scientific Reports*, 7: 1782. doi:10.1038/s41598-017-01831-z

Viswanathan, N., Dorsi, J. & George, S. (2014). The role of speechspecific properties of the background in the irrelevant sound effect. *Quarterly Journal of Experimental Psychology*, 67(3), 581–589.