Downloaded from orbit.dtu.dk on: Apr 23, 2024

DTU Library

Revisiting perceptual compensation for effects of reverberation in speech identifcation

Nielsen, Jens Bo; Dau, Torsten

Published in:
Acoustical Society of America. Journal

Link to article, DOI:
10.1121/1.3494508

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Nielsen, J. B., & Dau, T. (2010). Revisiting perceptual compensation for effects of reverberation in speech identifcation. *Acoustical Society of America. Journal*, *128*, 3088-3094. https://doi.org/10.1121/1.3494508

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Revisiting perceptual compensation for effects of reverberation in speech identification

Jens Bo Nielsen and Torsten Dau[a)]
Centre for Applied Hearing Research, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 1 July 2009; revised 6 September 2010; accepted 8 September 2010)

Listeners were given the task to identify the stop-consonant [t] in the test-word "stir" when the word was embedded in a carrier sentence. Reverberation was added to the test-word, but not to the carrier, and the ability to identify the [t] decreased because the amplitude modulations associated with the [t] were smeared. When a similar amount of reverberation was also added to the carrier sentence, the listeners' ability to identify the stop-consonant was restored. This phenomenon has in previous research been considered as evidence for an extrinsic compensation mechanism for reverberation in the human auditory system [Watkins (2005). J. Acoust. Soc. Am. **118**, 249–262]. In the present study, the reverberant test-word was embedded in additional non-reverberant carriers, such as white noise, speech-shaped noise and amplitude modulated noise. In addition, a reference condition was included where the test-word was presented in isolation, i.e., without any carrier stimulus. In all of these conditions, the ability to identify the stop-consonant [t] was enhanced relative to the condition using the non-reverberant speech carrier. The results suggest that the non-reverberant speech carrier produces an interference effect that impedes the identification of the stop-consonant. These findings raise doubts about the existence of the compensation mechanism.
© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3494508]

PACS number(s): 43.71.An, 43.66.Dc, 43.71.Bp, 43.71.Gv [MW] Pages: 3088–3094

I. INTRODUCTION

Reverberation can have a significant impact on the intelligibility of speech. Apart from background noise, reverberation is the environmental factor most often responsible for poor intelligibility. Reverberation smears the amplitude modulations of a sound signal such that energy peaks are prolonged and become less pronounced while energy dips are masked by preceding sounds (e.g., Houtgast and Steeneken, 1985). The attenuation of amplitude modulations in a room is an accurate predictor of speech intelligibility, and this correlation is reflected in methods for assessing intelligibility, such as the speech transmission index (STI; Steeneken and Houtgast, 1980). In a series of studies (e.g., Watkins, 2005c, 2005b; Watkins and Makin, 2007a, 2007b), Watkins considered the reduction of amplitude modulations in investigations of the impact of reverberation on speech intelligibility. He tested the ability of listeners to identify the stop-consonant [t] in the test-word "stir" when the word was embedded in a carrier sentence and different amounts of reverberation were added. When listeners could not identify the [t], the test-word was perceived as "sir." This degrading effect of reverberation can be assigned to both "overlap masking," which causes the [t] to be masked by energy from the preceding [s], and to "self-masking," caused by a smearing of energy within the [t] itself (Nábělek *et al.*, 1989). Stop-consonants are especially sensitive to reverberation because their identification depends on a rapidly changing sound amplitude that is easily smeared (e.g., Drullman *et al.*, 1994). Rather than focusing on the negative influence of reverberation, Watkins proposed that the human auditory system can, to a large extent, *compensate* for the effects of reverberation. His experiments demonstrated that when reverberation was added to the test-word, but not to the carrier sentence, more words were identified as "sir" because the modulations associated with the [t] in "stir" were smeared. However, when a similar amount of reverberation was also added to the carrier sentence, the number of "stir" identifications increased again. Watkins referred to this effect as *perceptual compensation* because he assumed that it was based on an enhanced ability of the listener to perceptually differentiate between the direct sound and the reverberation (Watkins, 2005c). Watkins also characterized the effect as *extrinsic* because it depends on information about the reverberation from surrounding speech and not the word itself. In his studies, Watkins investigated how the proposed compensation mechanism may pick up reverberation information from the context in order to increase speech intelligibility (e.g., Watkins, 2005c, 2005b; Watkins and Makin, 2007a, 2007b). The "tails" that reverberation adds to the offsets of sounds seem to be particularly important. He considered these tails to inform the compensation mechanism about the level of reverberation and potentially also to provide more detailed information about the acoustical environment.

The key observation in Watkins' investigations of the compensation mechanism was the shift in the listeners' perception of the reverberant test-word when embedded in two different speech carriers: (i) an almost non-reverberant carrier, and (ii) a carrier with a reverberation that matches the reverberation of the test-word. The interpretation of this shift as evidence for extrinsic compensation requires that the change in carrier reverberation per se causes the shift. How-

[a)]Author to whom correspondence should be addressed. Electronic mail: tdau@elektro.dtu.dk

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

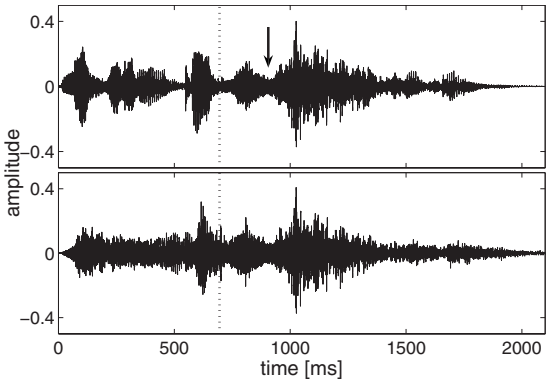

FIG. 1. Two examples of stimuli used in the study of Watkins (2005c). Both panels show the temporal waveforms of the sentence "next, you'll get… stir… to click on." In the upper panel, the carrier is slightly reverberant while the test-word is strongly reverberant. In the lower panel, the reverberation of the carrier was raised to the same level as for the test-word. The modulations of the carrier envelopes are more pronounced in the upper panel than in the lower panel. The vertical line indicates the transition between the preceding carrier and the test-word. The arrow points to the amplitude modulation dip originating from the [t] in "stir."

ever, adding reverberation to a speech signal changes several of the acoustical properties of the signal. Modulation depths are reduced and amplitude fluctuations important for phonetic identification are smeared. Specifically, it is possible that the shift in the perception of the test-word is caused by an interfering affect due to the presence of the non-reverberant carrier, instead of a compensating effect due to the reverberant carrier. Such an interfering effect could impede the identification of the [t] in the reverberant "stir." Since Watkins' experiments only considered the *difference* between listeners' perception of the test-word in combination with the reverberant versus the non-reverberant carrier and did not include any reference condition, it is not possible to decide which of the two carriers actually caused the perceptual shift. The alternative interpretation of the shift as a result of an interfering effect caused by the non-reverberant carrier was not considered in the studies by Watkins.

Forward masking in the amplitude modulation domain, potentially caused by modulation adaptation, could produce an interfering effect as the one described. Amplitude modulations inherent in a stimulus can mask modulations of similar rate in a subsequently presented stimulus whereby the amount of masking increases with increasing modulation depth of the masking stimulus (e.g., Wojtczak and Viemeister, 2005). The ability to identify the [t] in "stir" might thus decrease when the test-word is preceded by the more strongly modulated (non-reverberant) carrier than by the less modulated (reverberant) carrier. As a consequence, a listener might make more "stir" judgments with the reverberant carrier than with the non-reverberant carrier. In Fig. 1, Watkins' original carrier sentence "next, you'll get … to click on" is shown with the test-word "stir" embedded. In both panels, the part comprising "stir" is highly reverberant, but the distinctive modulation dip associated with the stop-consonant [t] is still visible (indicated by the arrow). In the upper panel, the carrier is essentially non-reverberant whereas carrier and test-word reverberations are matched in the lower panel. The amplitude modulations of the non-reverberant carrier (upper panel) are clearly more pronounced than those inherent in the reverberant carrier (lower panel).

In order to test whether an extrinsic compensation mechanism for reverberation is needed to account for the results observed in Watkins' studies, two experiments were conducted in the present study. In the first experiment, Watkins' original speech stimuli were used and some of the conditions from the original study were considered in a generalized experimental setup. In addition, a condition was included where the test-word was embedded in a (non-reverberant) white-noise carrier. If this non-reverberant noise carrier and the reverberant speech carrier would produce similar effects on the perception of the test-word, this would indicate that the shift in perception is probably not caused by a compensation effect for reverberation per se. In the second experiment, additional conditions with non-reverberant carriers were tested, including a speech-shaped noise, an amplitude-modulated noise and a reference condition with a silent carrier, i.e., a pause, in order to further examine the effect of the acoustical context on the listeners' perception of the test-word.

II. METHODS

A. Experimental procedure

The same experimental procedure as in Watkins (2005c) was used. In that study, binaural room impulse responses (BRIRs) were recorded at different distances from a sound source; the extreme distances were 0.32 m (here referred to as "near") and 10 m (here referred to as "far"). Dry speech recordings were convolved with these BRIRs in order to achieve test-word and carrier stimuli with different levels of reverberation. The influence of reverberation on the identification of the stop-consonant [t] was assessed using a continuum of 11 words, changing in steps from plain "sir" (step 0) to plain "stir" (step 10). The words were generated by calculating the envelopes of "sir" and "stir" and imposing different ratios of these envelopes on the waveform of "sir" (refer to Watkins, 2005c, for details). The test-words of the continuum were perceived as "sir" at low step numbers and as "stir" at high step numbers. The listeners switched from identifying the test-word as "sir" to "stir" at the so-called *category boundary*. The transition typically took place within a few continuum steps. In each experiment, a listener's category boundary was calculated as the total number of "sir" responses divided by 3 minus 0.5. The boundary thus quantified the listeners' ability to identify the [t] in the test-word. Watkins (2005c) showed that the boundary depended on the amount of reverberation added to the test-word and used the boundary as a measure of the degrading effect of reverberation. A high category boundary indicated that it was difficult for the listener to identify to [t] in "stir" and thus indicated a poorer speech intelligibility.

In the present study, the 11 continuum words were presented in combination with some of the speech carriers used in Watkins (2005c) as well as some additional carriers. All stimulus combinations were repeated three times.

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

B. Speech stimuli

The speech stimuli were based on Watkins' original carrier ("next you'll get … to click on") and his original "sir" to "stir" continuum. The original sound files consisted of the carrier and the embedded test-word in three combinations of reverberation: (i) "near" carrier and "near" test-word; (ii) "near" carrier and "far" test-word; (iii) "far" carrier and "far" test-word. For each combination, the carrier was combined with all 11 test-words from the continuum (from plain "sir" to plain "stir") resulting in 33 different sound files. The "near" reverberation produced a sound that was relatively dry with hardly noticeable reflections. With the "far" reverberation, the reflections were clearly noticeable, but the intelligibility of clear speech was not affected. Watkins used fast as well as slowly spoken stimuli in his investigations; the stimuli supplied by Watkins represented the slow version.

The experiments in the present study required that the test-words were combined with other carriers than the speech carrier used in Watkins (2005c). The original stimuli were therefore separated into individual waveforms. First, the 11 "near" sound files were divided into a "near" start carrier waveform ("next you'll get …"), a "near" end carrier waveform ("… to click on") and 11 different "near" test-word waveforms, in total 13 different waveforms. The reverberation time of the "near" waveforms was sufficiently short for the reverberation tails to be inaudible in the subsequent (and now separated) part of the signal. At the separation points, the onsets and offsets of the stimuli were ramped over a 10-ms interval by a squared cosine function. Second, one of the original "far" version waveforms (110 400 samples) was deconvolved with a corresponding "near" version (zero-padded to the same sample length as the "far" waveform) in order to create a filter function. The deconvolution was done by calculating the complex fast Fourier transform of both waveforms (110 400 points), dividing the "far" spectrum by the "near" spectrum, and applying the inverse fast Fourier transform to the result. Third, the 13 "near" waveforms were convolved with the filter function to create corresponding "far" versions of the stimuli. The available stimuli now consisted of "near" and "far" versions of the start carrier, the end carrier, and all 11 continuum steps of the test-word.[1] In the reverberation tails of the "far" waveforms, a faint tonal component could be heard as an artifact from the convolution. In order to investigate the impact of this, the convolved versions of the start carrier, the test-word, and the end carrier were recombined to create waveforms similar to the original "far" waveforms. The onsets of the three waveform parts were time-aligned with the original stimuli. The reverberation tail of the start carrier was thus superimposed with the test-word, and the tail of the test-word was superimposed with the end carrier. Informal listening tests revealed that the recombined waveforms were hardly distinguishable from the original waveforms.

C. Apparatus and procedure

The experiments were conducted in a double-walled sound insulated booth with a screen, a keyboard, and a mouse connected to an external PC. The stimuli were presented over Sennheiser HD580 headphones at an average sound pressure level of 65 dB SPL. A PC-application played the stimuli and waited for the listener to respond either "sir" or "stir" by a mouse click on one of two clearly marked buttons on the PC screen. The listener could also choose to respond by pressing "1" or "2" on the keyboard. After the response from the listener, a silent interval of 3 s duration was provided before the next stimulus. The experiments were preceded by a short training sequence that included examples of all the carriers that would be presented in the experiment.

III. EXPERIMENT 1: "SIR" VERSUS "STIR" IDENTIFICATIONS USING THE ORIGINAL STIMULUS SETUP

A. Rationale

This experiment was conducted to reproduce the key finding of Watkins (2005c) that the number of "stir" identifications increases when the amount of reverberation added to the carrier is increased. In addition, a white noise was included as a carrier that did not contain any reverberation information. Thus, in terms of compensation for reverberation, this carrier should not affect the listener's perception of the test-word.

B. Method

1. Listeners

Six listeners participated in the experiment. Participation was approved by the ethics committee of Copenhagen County. The listeners were aged between 24 to 42 and did not report any hearing problems. They were all students or employees of the Technical University of Denmark (DTU) and had previous experience with psychoacoustic experiments. The second author participated in the experiment. The listeners were fluent English speakers.

2. Stimuli

Watkins' 11-step "sir" to "stir" continuum was presented in three combinations with the "next you'll get… to click on" carrier: (i) "near carrier—near test-word," (ii) "near carrier—far test-word," and (iii) "far carrier—far test-word." The PC application that ran the experiment combined the separated carriers and test-words to recreate Watkins' original stimuli. When a reverberation tail was present in the stimulus, the tail was superimposed with the part of the stimulus that followed. The "far" version of the test-word was also presented in combination with a white-noise carrier. The white noise was presented at the same sound pressure level as the speech carrier (calculated separately for the start and the end portions of the carrier). The onsets of the test-word and the end carrier were time-aligned with Watkins' original speech stimuli. All stimuli were presented diotically in a different random order for each listener and repeated three times. In total, 4 stimulus combinations times 11 continuum steps times 3 repetitions were considered, resulting in 132 presentations.

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

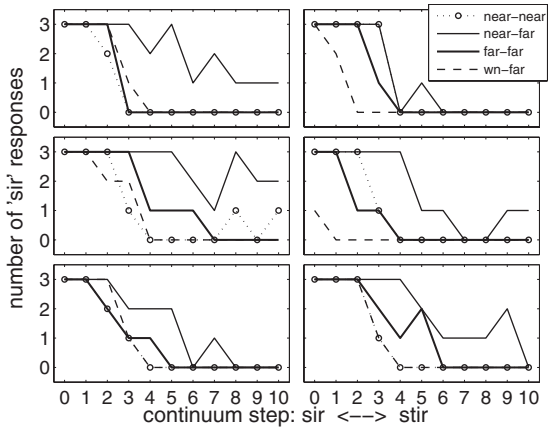

FIG. 2. Individual results from experiment 1 for the six listeners. The abscissa indicates the continuum steps of the test-word. The ordinate represents the number of "sir" responses obtained from three repetitions of each stimuli. For the listener shown in the upper left panel, the category boundary was 2.2 for the near-carrier near-test-word ("near-near") condition, 2.5 for the "far-far" condition, 2.8 for the "wn-far" condition and 7.2 for the "near-far" condition.

C. Results and discussion

Figure 2 shows the individual results obtained in experiment 1 for the six listeners. For three of the carrier/test-word combinations ("near—near," "far—far," and "wn—far"), the transition from "sir" to "stir" responses typically occurred within 2–3 steps, while the transition was less consistent for the "near—far" combination. Nevertheless, the category boundary was calculated for all combinations. For example, the boundary of the "near—far" combination in the upper left panel was calculated as $(4\cdot 3+2+3+1+2+3\cdot 1)/3-0.5=7.2$, using the calculation procedure from Watkins (2005c). For all listeners, the "near—far" combination (thin solid curve) resulted in a transition from "sir" to "stir" at a higher step number than for the other combinations, indicating that the identification of the [t] in "stir" was most difficult for this combination.

The average category boundaries across the six listeners are shown in Fig. 3 as a function of the stimulus condition. The boundary in the "near—near" condition was found to be at a value of 2.9. Adding reverberation to the test-word ("near—far" condition) shifted the boundary to 6.2, while adding reverberation also to the carrier ("far—far" condition) shifted it to 3.2, a value close to the boundary obtained for the "near—near" condition. The "wn—far" boundary was found to be at an even lower step number of 2.1. A one-way ANOVA followed by a Tukey-Kramer post-hoc analysis showed that there were no significant differences between the three low-level boundaries, but a highly significant difference between each of these and the "near—far" boundary [$p<0.005$]. The observed category boundaries are in agreement with the results of Watkins (2005c) obtained in a similar experiment (see his "slow, L-shaped" condition of experiment 1, p. 253). For direct comparison, the corresponding (mean) results from that study are indicated as crosses in Fig. 3.

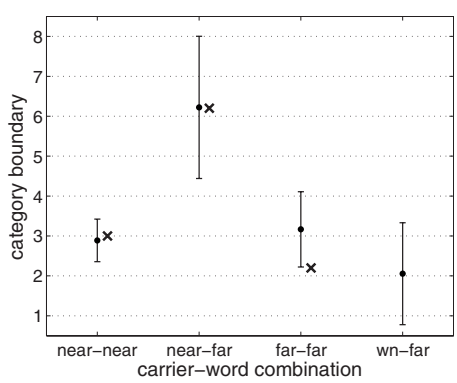

FIG. 3. Mean category boundaries for the six listeners in experiment 1. Bars indicate one standard deviation. The boundary for the "near—far" condition is significantly higher than for the remaining conditions. The results from a similar experiment in Watkins (2005c) (experiment 1, "slow, L-shaped" condition) are marked by crosses.

Thus, experiment 1 confirms the observations in Watkins (2005c). When applied to the test-word, the "far" reverberation significantly increases the category boundary because it becomes more difficult to identify the [t] in "stir." When this reverberation is also applied to the carrier, the category boundary shifts back and has a value close to that obtained in the "near—near" condition. Thus, the carrier reverberation facilitates the identification of the [t] which, in principle, might be a consequence of the proposed compensation mechanism for reverberation. However, this interpretation appears inconsistent with the low category boundary observed for the "wn—far" condition. The boundary shift between the "near" and the "white-noise" carrier is difficult to explain in terms of a compensation effect related to reverberation since the white noise does not contain reverberation information. Specifically, the white noise does not contain any of the reverberation tails that have been assumed to be particularly important for compensation (e.g., Watkins, 2005c, 2005b). Instead, this result suggests that the listeners' perception of the "sir" to "stir" continuum could be affected by carrier properties that are not necessarily associated with reverberation. The "near" carrier seems to provide an interference with the listeners' perception of the "far" test-word such that a significantly higher category boundary is observed.

IV. EXPERIMENT 2: EFFECTS OF OTHER NON-REVERBERANT CARRIERS ON "SIR" VERSUS "STIR" IDENTIFICATIONS

A. Rationale

The second experiment was performed to further investigate the effect of the carrier stimulus properties on the category boundary. A silent interval was included as one of the carriers and regarded as a "reference" condition. In addition, two other non-reverberant carriers were considered, a modulated noise and an unmodulated speech-shaped noise.

B. Listeners

The experiment was conducted with 13 listeners, aged between 26 and 43 years, who did not report any hearing problems. They were all students or employees of DTU and had previous experience in psychoacoustic experiments. The

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

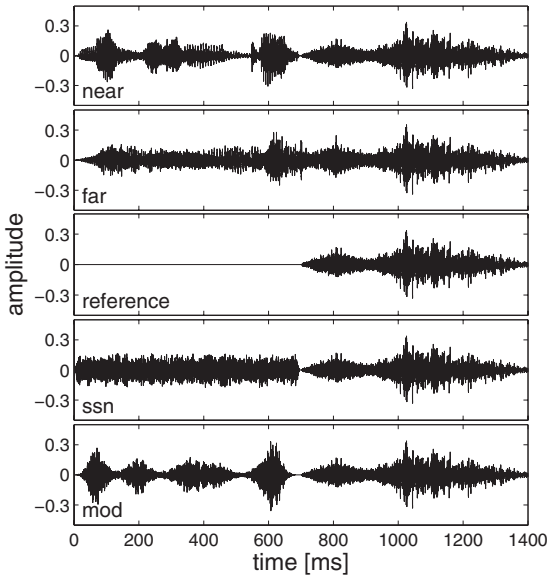

FIG. 4. Carrier stimuli used in experiment 2. The test-word was not followed by a carrier component as in experiment 1. In the figure, all carriers are shown with the "far" version of step 10 of the test-word continuum. The stimuli in the two upper panels were created from the separated versions of Watkins' original speech stimuli. The carrier in the third panel is a silent interval and represents the reference condition. The fourth and fifth panel show unmodulated speech shaped noise ("ssn") and modulated speech-shaped noise ("mod") as the carrier.

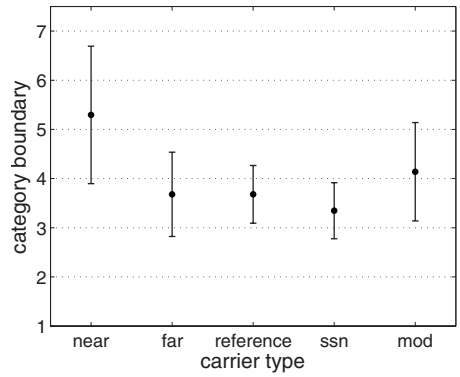

FIG. 5. Mean category boundaries obtained in experiment 2. Bars indicate one standard deviation. The boundary for the "near" carrier is significantly higher than the boundary for the three middle carriers. The boundary for the modulated noise carrier ("mod") is significantly higher than the boundary for the unmodulated noise ("ssn").

second author participated in the experiment. All listeners were fluent English speakers. Participation was approved by the ethics committee of Copenhagen County.

C. Stimuli

Five different carriers were presented in combination with the "far" version of the "sir" to "stir" continuum. The "near" version of the test-word was not included since the evidence of the proposed compensation mechanism relates only to the "far" test-word in different contexts. Figure 4 shows the different stimulus waveforms including step 10 of the "far" test-word. The part of the original carrier that followed the test-word ("… to click on") was omitted in this experiment since it did not have a meaningful correlate for the non-speech carriers. According to Watkins (2005a), this omission should not affect the perception of the test-word. The five carriers included in the experiment were: The first part of the original "near" carrier (top panel), the first part of the original "far" carrier (second panel), a silent pause of the same duration as the original speech carrier, reflecting a "reference" condition (third panel), an unmodulated speech-shaped noise with a magnitude spectrum corresponding to that of the talker of the original speech material (fourth panel) and, finally, the same speech-shaped noise imposed by random amplitude modulations in the range from 4 to 8 Hz. The unmodulated speech-shaped noise was produced by superimposing Watkins' original "near" speech material 150 times with randomly shifted offsets, using a similar procedure as in Wagener *et al.* (2003). The modulated noise was produced by multiplying the speech-shaped noise with a modulator containing random frequencies between 4 and 8 Hz and modulation depth 1. The average sound pressure level of the different carriers, except the reference carrier, was 65 dB SPL. The onset of the test-word was time-aligned with Watkins' original speech stimuli. All stimuli were presented diotically. The total number of presentations per listener was 165, resulting from 5 carriers times 11 continuum steps times 3 repetitions. The presentations were done in a different random order for each listener.

D. Results and discussion

Figure 5 shows the category boundaries, averaged across listeners, for the five different carriers that were combined with the "far" test-word. The category boundary was 5.3 for the original "near" carrier, 3.7 for the original "far" carrier, 3.7 for the "reference" carrier, 3.3 for the "ssn" carrier and 4.1 for the "mod" carrier. A one-way ANOVA followed by a Tukey-Kramer post-hoc analysis showed no significant differences between the four low-level boundaries, but a significant difference between each of these and the "near—far" boundary [$p<0.01$]. The "near" speech carrier represents the only condition that causes a significantly different boundary level; this carrier thus appears to have an exceptional effect on the listeners' perception of the test-word. The low category boundary obtained in the reference condition and the speech-shaped noise condition ("ssn") cannot be explained by a compensation effect related to reverberation. When comparing the boundaries obtained with the unmodulated and the modulated noise ("ssn" and "mod"), a t-test provided a significant difference between the two boundaries [$p=0.03$]. These two carriers differ with respect to their amplitude modulations (Fig. 4), while none of them are associated with reverberation. This suggests that the listeners' category boundary for the test-word is affected by the content of amplitude modulation energy in the carrier. As shown in Fig. 1, also the original "near" and "far" speech carriers differ with respect to their degree of modulation. It is thus likely that at least part of the boundary shift between these two carriers is due to a change in their amplitude modulation content.

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

V. GENERAL DISCUSSION

The proposed extrinsic compensation effect in Watkins (2005c) was based on the observation of an increased number of "stir" identifications in a "sir" to "stir" continuum when the reverberant test-word was embedded in a speech carrier with the same amount of reverberation. In the experiments of the present study, a similar increase of "stir" identifications was found with other carriers that did not contain any information related to the reverberation of the test-word. In fact, all other carriers tested in this study, including the reference condition, led to lower category boundaries than the original "near" speech carrier. This suggests that the boundary shifts between the "near" and the "far" speech carrier conditions observed in both Watkins (2005c) and the present study were not caused by a reverberation compensation effect. More likely, the "near" carrier produced an interfering effect which led to a reduction of the number of [t] identifications in "stir."

The higher boundary that is obtained for the modulated noise carrier in comparison to the unmodulated noise (condition "mod" and "ssn" in Fig. 5) suggests that the proposed interfering effect of the non-reverberant "near" carrier is related to its modulation content. The increased amount of amplitude modulation energy in the speech relevant modulation frequency range (≈4–20 Hz) might have produced a forward masking effect on the modulations in the test-word. Wojtczak and Viemeister (2005) measured amplitude modulation forward masking in a modulation detection task and found that the sensitivity to amplitude modulations can be substantially lower even after a brief exposure to modulations of a similar rate. The duration of the (non-speech) masker in Wojtczak and Viemeister's experiments was 150 ms. They observed an exponential recovery from the masking effect and a threshold that remained elevated for at least 150 ms. This is approximately the time interval between the offset of the "near" carrier and the position of the [t] in "stir." A forward masking effect in the amplitude modulation domain, stemming from the relatively strong modulations contained in the "near" carrier, may therefore be responsible for at least part of the interfering effect produced by this carrier. However, since modulation forward masking effects are highly audio-frequency and modulation-frequency specific (e.g., Kay and Matthews, 1972), an amplitude modulated broadband noise cannot be expected to produce the same amount of masking as speech, which consists of both complex amplitude as well as frequency modulations that vary over time. Specifically, a carrier containing dry speech of the same speaker is likely to produce maximal interference when attempting to identify the amplitude modulation of the [t] in "stir."

The assumption of an interfering effect produced by the "near" carrier as an alternative explanation for the boundary shifts observed in Watkins (2005c) appears to be supported by some of the original experimental results found in that same study. In experiment 2 of Watkins (2005c), the reverberation that was added to the test-word and the speech carrier, respectively, originated from two different locations, an L-shaped room and a corridor. The results were compared to an experimental setup with reverberation from the *same* location applied to the carrier and the test-word (Fig. 3 in Watkins, 2005c). The "compensation effect" turned out to be substantial even when there was a switch in the location between the carrier and the test-word. Watkins concluded that the effect was independent of details in the carrier reverberation and argued that the effect seemed to rely on aspects of the reverberation that were common in the two locations. However, a more straightforward explanation might be provided based on an interfering effect of the "near" carrier. The "near" BRIRs of the L-shaped room and the corridor are similar because the direct sound dominates in the impulse responses. The two "near" speech carriers are thus similar and produce a similar interference effect.

In the same study, Watkins investigated the effect of dichotic versus monaural presentation of the stimuli. The "compensation for reverberation" effect was found to be *greater* in the monaural condition than in the dichotic condition (see Fig. 3 in Watkins, 2005c). This result appears surprising since a dichotic signal typically contains more information about the reverberation in a room than a monaural signal and since binaural listening (relative to monaural listening) typically improves speech intelligibility in reverberant conditions (e.g., Nábělek and Robinson, 1982). Watkins (2005c) concluded that extrinsic compensation does not use binaural information but stems from a monaural mechanism. He also argued that the binaural benefit in reverberant conditions is "intrinsic" and thus independent of the extrinsic carrier reverberation. He proposed two distinctly different ways in which the auditory system might benefit from reverberation information: an analysis of "intrinsic" reverberation based on binaural information and an analysis of "extrinsic" reverberation based on monaural information. Such a distinction would not be required if the results were explained in terms of an interfering effect. Since binaural listening generally improves consonant perception in reverberation compared to monaural listening (Helfer, 1994), the perception of the test-word would be more affected by the interfering effect in the case of monaural presentation than in the case of binaural presentation. This should produce greater boundary shifts in the monaural condition than in the binaural condition. This interpretation also seems supported by the results from investigations with various noise carriers in Watkins (2005c) and Watkins and Makin (2007a, 2007b). These investigations included a broadband noise carrier that had the temporal envelope of the original dry speech carrier. When reverberation was added to this noise carrier, a boundary shift similar to that for the speech carrier was observed for monaural stimulus presentations (Watkins and Makin, 2007a, 2007b), but not for dichotic stimuli (Watkins, 2005c). In terms of an interfering effect, the modulated (non-reverberant) noise seems to degrade consonant perception in the monaural case, but the effect is opposed by binaural processing.

Finally, the study of Watkins (2005c) also included measurements of the effect of carrier reverberation on "near" test-words that only contained a very small amount of reverberation. These conditions were not considered in the present study. The results showed that the category boundary obtained for the "near" words depended on the carrier rever-

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

beration in a similar way as for the "far" test-words. The boundary was found to be lower when the test-word was combined with the "far" carrier than when it was combined with the "near" carrier (Figs. 2, 3, and 4 in Watkins, 2005c). There was a floor-effect restricting the size of the boundary shift for the "near" test-word; nevertheless, the same direction of the shift was observed in all experiments. Watkins' data thus indicated that the "far" speech carrier *always* leads to more "stir" identifications than the "near" carrier, even when the carrier is misleading with respect to the reverberation of the test-word. This further suggests that the shifts of the category boundary are not related to a compensation mechanism that is informed by the reverberation in the carrier, but to an interfering effect of the "near" carrier, potentially caused by modulation forward masking.

VI. CONCLUSION

This study investigated the compensation for reverberation hypothesis proposed by Watkins (2005c), which states that the human auditory system can perceptually compensate for the negative effect of reverberation on speech intelligibility. The compensation mechanism is assumed to require "reverberation information" from preceding speech to be operational and the effect has therefore also been termed "extrinsic compensation."

The results of the present study are difficult to explain within the concept of this compensation mechanism since most conditions with non-reverberant carriers, including a silent interval representing the reference condition, produced the same results as the reverberant speech carrier. Rather, the results seem consistent with an interfering effect produced by the specific stimulus characteristics of the "near" speech carrier. This carrier produced a category boundary that was distinctly higher than any other tested carrier. The original data of Watkins (2005c) are consistent with an interpretation based on such an interfering effect. It is suggested that the interference is, at least partly, caused by modulation forward masking produced by the envelope fluctuations inherent in the carrier. However, additional investigations, including quantitative modeling work, are needed to further evaluate this hypothesis.

ACKNOWLEDGMENTS

The authors wish to thank students and colleagues at the Centre for Applied Hearing Research for participation in the experiments, and A. J. Watkins for supplying his original speech stimuli. The authors would also like to thank Associate Editor M. Wojtczak and two anonymous reviewers for their comments and suggestions. The present work was supported by the Oticon Foundation.

[1]The "near" and "far" versions of the speech carriers and the test-word (steps 0 and 10) are available as wav-files (see Supplemental Material). These separated waveforms have the prefix "S." The stimuli for experiment 1 (in combination with step 0 of the test-word) and for experiment 2 (in combination with step 10 of the test-word) are also available.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–64.

Helfer, K. S. (**1994**). "Binaural cues and consonant perception in reverberation and noise," J. Speech Hear. Res. **37**, 429–438.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

Kay, R. H., and Matthews, D. R. (**1972**). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," J. Physiol. (London) **225**, 657–677.

Nábělek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**, 1259–1265.

Nábělek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.

See Supplemental Material at http://dx.doi.org/10.1121/1.3494508 Document No. E-JASMAN-128-053011. Files containing examples of the stimuli used in the study.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.

Watkins, A. J. (**2005a**). "Listening in real-room reverberation: Effects of extrinsic context," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 423–428.

Watkins, A. J. (**2005b**). "Perceptual compensation for effects of echo and of reverberation on speech identification," Acta. Acust. Acust. **91**, 892–901.

Watkins, A. J. (**2005c**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.

Watkins, A. J., and Makin, S. J. (**2007a**). "Perceptual compensation for reverberation in speech identification: Effects of single-band, multiple-band and wideband noise contexts," Acta. Acust. Acust. **93**, 403–410.

Watkins, A. J., and Makin, S. J. (**2007b**). "Steady-spectrum contexts and perceptual compensation for reverberation in speech identification," J. Acoust. Soc. Am. **121**, 257–266.

Wojtczak, M., and Viemeister, N. F. (**2005**). "Forward masking of amplitude modulation: Basic characteristics," J. Acoust. Soc. Am. **118**, 3198–3210.

Downloaded 08 Dec 2010 to 192.38.67.112. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp