

# Speech Recognition and Noise Adaptation in Realistic Noises

Trends in Hearing

Volume 29: 1–13

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/23312165251343457

journals.sagepub.com/home/tia



Miriam I. Marrufo-Pérez<sup>1,2</sup> and Enrique A. Lopez-Poveda<sup>1,2,3</sup>

## Abstract

The recognition of isolated words in noise improves as words are delayed from the noise onset. This phenomenon, known as adaptation to noise, has been mostly investigated using synthetic noises. The aim here was to investigate whether adaptation occurs for realistic noises and to what extent it depends on the spectrum and level fluctuations of the noise. Forty-nine different realistic and synthetic noises were analyzed and classified according to how much they fluctuated in level over time and how much their spectra differed from the speech spectrum. Six representative noises were chosen that covered the observed range of level fluctuations and spectral differences but could still mask speech. For the six noises, speech reception thresholds (SRTs) were measured for natural and tone-vocoded words delayed 50 (early condition) and 800 ms (late condition) from the noise onset. Adaptation was calculated as the SRT improvement in the late relative to the early condition. Twenty-two adults with normal hearing participated in the experiments. For natural words, adaptation was small overall (mean = 0.5 dB) and similar across the six noises. For vocoded words, significant adaptation occurred for all six noises (mean = 1.3 dB) and was not statistically different across noises. For the tested noises, the amount of adaptation was independent of the spectrum and level fluctuations of the noise. The results suggest that adaptation in speech recognition can occur in realistic noisy environments.

## Keywords

sound-level statistics, noise, speech intelligibility

Received: September 30, 2024; revised: April 29, 2025; accepted: May 1, 2025

## Introduction

In everyday life, people are surrounded by diverse types of noise that can hamper communication. To some extent, listeners can adapt to the noise as they gradually recognize more words when words are delayed relative to the onset of the noise (Ainsworth & Meyer, 1994). Except for one study, adaptation to noise in speech recognition has always been investigated using synthetic noises (e.g., Ainsworth & Meyer, 1994; Cervera & Ainsworth, 2005; Cervera & Gonzalez-Alvarez, 2007; Marrufo-Pérez et al., 2018). The exception study (Khalighinejad et al., 2019) was limited to three urban noises and did not investigate the importance of the noise characteristics for adaptation. One aim of the present study was to investigate whether adaptation to noise in word recognition occurs for realistic noises. A second aim was to investigate if adaptation depends on the similarity between the noise and speech spectra and/or on the noise-level fluctuations.

Normal-hearing (NH) listeners adapt to the noise background with a time course of about 350 ms (Ben-David et al., 2012, 2016). The mechanisms responsible for

adaptation to noise are unclear. A possible mechanism is neural dynamic range adaptation toward the most common level in the noise preceding the word (Ainsworth & Meyer, 1994; Marrufo-Pérez & Lopez-Poveda, 2022), a phenomenon known as “adaptation to noise level statistics.” This kind of adaptation likely facilitates the encoding of speech temporal (Marrufo-Pérez et al., 2018, 2020; Marrufo-Pérez & Lopez-Poveda, 2022) and spectral (Ainsworth & Meyer,

<sup>1</sup>Instituto de Neurociencias de Castilla y León (INCYL), Universidad de Salamanca, Salamanca, Spain

<sup>2</sup>Instituto de Investigación Biomédica de Salamanca (IBSAL), Universidad de Salamanca, Salamanca, Spain

<sup>3</sup>Departamento de Cirugía, Facultad de Medicina, Universidad de Salamanca, Salamanca, Spain

## Corresponding author:

Enrique A. Lopez-Poveda, Instituto de Neurociencias de Castilla y León, Universidad de Salamanca, Calle Pintor Fernando Gallego 1, 37007 Salamanca, Spain.

Email: ealopezpoveda@usal.es



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

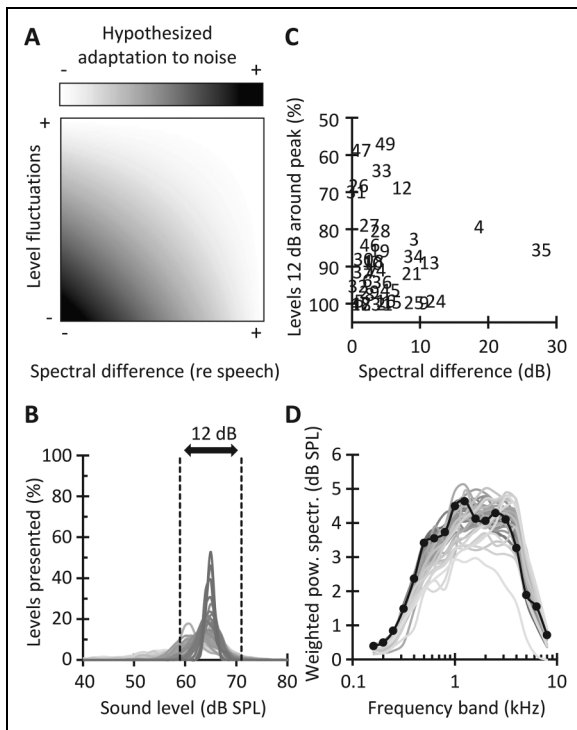
1994) cues for speech tokens in noise and possibly contributes to creating noise-tolerant speech representations at the auditory cortex (Khalighinejad et al., 2019; King & Walker, 2020; Rabinowitz et al., 2013).

If neural dynamic range adaptation to the noise-level statistics contributes to behavioral adaptation to noise, one would expect behavioral noise adaptation to be the largest for noises with minimal or no level fluctuations (steady noises). Physiological studies have revealed that when a noise level is changing every 50 ms and 80% of the levels are presented within a range of 12 dB (defined as the high-probability region), auditory neurons displace their dynamic range toward the high-probability region and encode the levels of that region more accurately than when the width of the high-probability region is 24 dB (Dean et al., 2005). In addition, when the distribution of noise levels has two high-probability regions, the levels presented at these regions are encoded less accurately than when the same sound levels are presented

within a single high-probability region (Dean et al., 2005). These findings suggest that the benefits of neural dynamic range adaptation to the noise-level statistics are larger when the level distribution is narrower. According to this, one would expect greater behavioral adaptation for realistic noises that are steady rather than fluctuating in level. There is some evidence supporting this assumption. For instance, Ben-David et al. (2012, 2016) revealed that adaptation was less for babble noise than for steady, speech-shaped noise (SSN). Also, Marrufo-Pérez et al. (2020) showed that for speech embedded in SSN, adaptation occurred when the precursor (or adaptor) noise was steady but did not occur when the precursor noise was fluctuating in level.

If neural dynamic range adaptation to the level statistics of the noise preceding speech contributes to behavioral adaptation to noise, behavioral adaptation should also depend on the overlap between the neural excitation patterns evoked by the noise and the speech; that is, the greater the number of neurons that respond to the noise and the speech, the greater the behavioral adaptation. Therefore, for a fixed signal-to-noise ratio (SNR), one would expect behavioral adaptation to gradually decrease with increasing the difference between the noise and speech spectra. Previous studies suggest that the noise spectrum is important for adaptation to noise (Cervera & Ainsworth, 2005; Cervera & Gonzalez-Alvarez, 2007). For instance, Cervera and Gonzalez-Alvarez (2007) presented bandpass-filtered syllables (0.92–2 kHz) and found more adaptation when the noise was filtered with the same bandpass filter as speech than when it was filtered with a narrower or a broader bandpass filter.

The importance of noise-level fluctuations or long-term spectrum for adaptation to noise has been investigated using synthetic noises as those described above. The relevance of the characteristics of daily life noises for adaptation, however, has received little attention. To our knowledge, only Khalighinejad et al. (2019) have measured noise adaptation in speech recognition using realistic noises. They measured the recognition of six phonemes (/pa/, /ta/, /ka/, /ba/, /da/, /ga/) presented early or late relative to the onset of three urban noises (“jet,” “city,” “bar”). Here, we extend their findings by measuring adaptation to noise in NH listeners presented with phonetically balanced disyllabic words belonging to everyday vocabulary. In addition, we analyzed 49 realistic and synthetic noises and classified them according to their level fluctuations and spectrum. The spectra were weighted according to the relative importance of the different frequency components for speech recognition. This classification allowed us to shed light on the importance of spectral and temporal components of realistic noises for adaptation, something that has not been investigated before. We hypothesized that the more fluctuating a (realistic) noise is, and the more different its spectrum is from the speech spectrum, the less the adaptation to noise will be (Figure 1A). We tested this hypothesis by measuring adaptation for six out of the 49 analyzed noises chosen to differ in level fluctuations but not in spectra and vice versa.



**Figure 1.** Hypothesis and Analyses of the Spectra and Level Fluctuations of 39 Noises. (A) Hypothesized adaptation to noise as a function of the noise level fluctuations and the difference between the speech and noise spectra. (B) Level distributions for 39 noises. To obtain the histogram, sound levels were obtained for nonoverlapping time windows of 50 ms. Each line depicts the distribution for one noise. (C) Noises classified according to their level fluctuations and their spectral difference relative to speech. Each number corresponds to one noise specified in Table 1. (D) Spectra of the natural speech (black) and the noises (gray) for different 1/3-octave frequency bands with center frequencies between 0.16 and 8 kHz. The spectra are weighted according to the importance of each frequency band for speech recognition.

## Material and Methods

### Noise Classification

To investigate if adaptation depends on the noise-level fluctuations and/or on the similarity between the noise and speech spectra, we analyzed the spectra and level fluctuations of 49 noise samples (listed in Table 1). Forty-six of them were realistic noises taken from The Natural Sound Library (Bjerg & Larsen, 2006), one was SSN (Nilsson et al., 1994), one was the International Female Fluctuating Masker (IFFM) (Holube, 2011), and one was the “hair dryer” noise from The Natural Sound Library modulated by the envelope of the IFFM (hereinafter modulated “hair dryer” noise). Speech-shaped noise and IFFM were included because they both had speech-like spectra but were steady and fluctuating, respectively. The modulated hair dryer noise was included because it was fluctuating but its long-term spectrum was like that of the hair dryer noise, and thus it allowed us to investigate the effect

of noise-level fluctuations independent from the noise spectrum.

The noise-level fluctuations were calculated using a procedure inspired by physiological studies of neural adaptation to sound-level statistics (e.g., Dean et al., 2005). We first set the noise level for the whole sound file at 65 dB SPL. Then, we calculated the level in nonoverlapping time windows of 50 ms in duration and created a histogram from the levels obtained in these windows. Lastly, we calculated the percentage of levels that were 12 dB around the peak in the distribution, as illustrated in Figure 1B. We found two noises (“Door closing” and “Zoo (Birds)”) whose level distributions had peaks well below 65 dB SPL because they had a constant silence (“Door closing”) or noise background (“Zoo”) with sudden high-level sounds. These two files were removed from the analyses and are not shown in Figure 1 (but are shown in Supplemental Figure S1).

To quantify the spectral difference between each noise and speech, we considered the importance of different spectral bands for speech recognition. The spectral bands and their importance were based on the Speech Intelligibility Index (SII) (Table 3 in ANSI/ASA S3.5–1997 (R2007)). The procedure was as follows. First, we applied a Fast Fourier Transform to obtain the long-term average spectrum of the noise. Then, we split the spectrum into the 18 frequency bands used to calculate the SII (the bands were approximately logarithmically spaced with center frequencies from 0.16 and 8 kHz). The squared amplitude for the frequency bins within a given band was summed, and the result was expressed in decibels (dB) by applying  $10 \times \log_{10}$ . The resulting value was multiplied by a weight (importance for speech recognition) obtained from the SII (Figure 1D). The same procedure was applied to a speech signal that we created by concatenating all the words. The same duration of the noise and speech files was analyzed (54.4 s, which corresponds to the duration of the shortest noise file). The difference between the speech and noise spectra was obtained by first squaring the difference (in dB) for each frequency band and then summing all the differences to obtain a single value (Figure 1C). This analysis revealed a group of noises whose spectra were so different from speech that they would require very large SNRs to yield 50% word recognition (or may not mask speech at all). Because here the speech reception threshold (SRT; the SNR at 50% word recognition) was measured using fixed-level noise (at 65 dB SPL) and varying the speech level (see below), it is likely that for those noises the SRT could reflect the speech audibility threshold rather than a masked threshold. To prevent this from happening, the noises in question (indicated with an asterisk in Table 1) were removed from the analyses and are not shown in Figure 1C and D (but are shown in Supplemental Figure S1).

After analyzing the level fluctuations and spectra of all noises we chose six of them to measure adaptation. The criteria behind our choice are explained in the “Results” section.

**Table 1.** Noises Analyzed.

01. Hair Dryer	26. Classic Music
02. Vacuum Cleaner	27. Soft Music
03. Circular Saw	28. Jazz Music
04. Angle Grinder	29. Rock Music
05. Cantina	30. Football Match (Stadium 6000–8000 People)
06. Traffic Noise (Low Intensity)	31. Choir In Church
07. Traffic Noise (High Intensity)	32. Choir In Church W. Organ
08. Lathe	33. Flute (Soft)
09. Ventilation	34. Pneumatic Drill
10. Industrial Dishwasher	35. Near Airport
11. Door closing*	36. Near Airplane
12. Keyboard Typing	37. Party (60 People)
13. Kitchen	38. Car slow in city*
14. Coffee Machine	39. Car 60km*
15. Bathwater	40. Car motorway (rough) *
16. Supermarket	41. Car motorway*
17. Shopping Centre	42. Car accelerating*
18. Pedestrians	43. Inside train (bumpy)*
19. Children Playing (Inside)	44. Inside train*
20. Zoo (Birds) *	45. Underground Station
21. Forest Birds (Very Soft)	46. Train Stopping Starting
22. Inside bus*	47. IFFM
23. Truck Idling	48. SSN
24. Building Site 1 (Small Wacker)	49. Modulated Hair Dryer
25. Building Site 2 (Big Wacker)	

The number indicated for each noise corresponds to that plotted in Figure 1C. The noises with asterisks are not plotted in Figure 1 (see main text for details).

## Participants

Twenty-two people (10 male) with NH participated in the experiments (mean age = 26.4 years; standard deviation [SD] = 5.9 years). The experiments involved the presentation of the stimuli to the two ears. Seventeen listeners had audiometric thresholds less than or equal to 20 dB hearing level (HL) in the two ears at octave frequencies between 250 Hz and 8 kHz (ANSI, 1996). Five participants had audiometric thresholds higher than 20 dB HL and lower than 35 dB HL at 8 kHz in one or two ears. All participants were native speakers of Spanish. They were volunteers and not paid for their time. All of them signed an informed consent to participate in the study. Methods were approved by the Ethics Committee of the University of Salamanca (Spain).

## Stimuli

Speech reception thresholds were measured for phonetically balanced disyllabic words belonging to everyday vocabulary (Cárdenas & Marrero, 1994). Words were unprocessed (hereinafter referred to as “natural”) or processed to maintain the envelope cues and disregard temporal fine structure (TFS) cues (hereinafter referred to as “vocoded”). Vocoded words were included because they usually result in larger adaptation to noise than natural words and thus the effect of noise characteristics may be better shown with these words (Marrufo-Pérez et al., 2018, 2020). The vocoder was applied to the words only (not to the noise) as we wanted to investigate the effect of removing particular speech cues on adaptation. The vocoder included a high-pass preemphasis filter (first-order Butterworth filter with a 3-dB cutoff frequency of 1.2 kHz); a bank of 12, sixth-order Butterworth bandpass filters whose 3-dB cutoff frequencies followed a modified logarithmic distribution between 100 and 8500 Hz; and envelope extraction via full-wave rectification and low-pass filtering (fourth-order Butterworth low-pass filter with a 3-dB cutoff frequency of 400 Hz). The envelope for each frequency channel was used to modulate a sinusoidal carrier at the channel center frequency, and the modulated signals were filtered again through the corresponding filter in the bank, and sample-wise added to obtain the vocoded speech.

For each of the six chosen noises, SRTs were measured in two conditions referred to as “early” and “late.” In the early condition, words were delayed 50 ms from the onset of the noise. In the late condition, words were delayed 800 ms from the onset of the noise. The SRT improvement in the late relative to the early condition was regarded as the amount of adaptation to noise. The noise always finished 50 ms after the word offset and included raised-cosine onset and offset ramps of 50 ms. Different noise segments were used for each word presentation, that is, noises were not frozen.

To assess adaptation in realistic scenarios, stimuli (speech and noise) were presented to both ears. The noises from The Natural Sound Library were recorded using a Brüel & Kjaer

artificial head and torso simulator with two microphones positioned at the entrances of the ear canals. This ensured that the noises included natural spatial location cues induced by the head of the manikin. In our experiment, we fixed the level of the signal in the left ear at 65 dB SPL, and the level in the right ear varied according to natural interaural level differences. The speech, SSN and IFFM sound files, however, were monophonic and did not contain spatial cues. To make the presentation of these stimuli more realistic, they were filtered with head-related transfer functions (HRTFs) for an acoustic KEMAR manikin (Gardner & Martin, 1995). The word signals were HRTF-filtered to simulate that they were presented at 0° azimuth and 0° elevation. The SSN and IFFM signals were filtered to simulate sources located at −5° azimuth and 0° elevation as it would be unrealistic that the noise source was colocated with the speech source. The levels of the SSN and IFFM noises were 65 dB SPL prior to spatial processing (64 dB SPL in the left ear after spatial processing). It is unlikely that the use of two different manikins (Brüel & Kjaer vs. KEMAR) influenced the results as they both produce similar spatial cues (see, e.g., Snaidero et al., 2011).

A sound cue (1-kHz pure tone with 500 ms duration and 66 dB SPL) was presented 500 ms before the noise onset to warn the listener about the stimulus presentation and to focus his/her attention on the speech recognition task. Without the cue, the listener may have been more distracted in the early condition than in the late condition (because the noise served as a cue in the late condition), which may have produced a fake temporal effect (Marrufo-Pérez et al., 2018).

## Procedure

Twenty-five disyllabic words (or trials) were used to measure each SRT. They corresponded to one of the 10 phonetically balanced lists from Cárdenas and Marrero (1994). Words from a list were presented in random order across test conditions to minimize the possibility that participants remembered the words. Each list was presented eight times to each participant. To measure an SRT, the noise level was fixed (at 65 dB SPL in the left ear; nonweighted), and the speech level varied adaptively using a one-down, one-up adaptive rule, that is, the speech level decreased after a correct response and increased after an incorrect response. The SRT was thus defined as the SNR giving 50% correct word recognition in the psychometric function (Levitt, 1971). The initial SNR was −5 dB for natural words and 0 dB for vocoded words for all noises except for the “angle grinder” noise, for which the initial SNR was −10 dB and −5 dB for natural and vocoded words, respectively (the initial SNRs were lower for the “angle grinder” noise because pilot tests revealed that the SRTs were more negative for this noise than for the other noises and we wanted the adaptive procedure to converge at approximately the same number of trials for all noises). The speech level changed in 3-dB steps between words/trials 1 and 5, and in 1-dB steps

between words/trials 5 and 25. The SRT was calculated as the mean of the SNRs for the final 18 words/trials (the SNR for the 19th word/trial was calculated and used in the SRT estimate but not actually presented). Feedback was not given to the participants on the correctness of their responses.

Speech reception thresholds were measured in 12 conditions (6 noises  $\times$  2 temporal positions [early and late]) for the two-word types (natural and vocoded). An SRT measurement was discarded, and a new SRT was then measured immediately after when the SD within the measure was higher than 3.5 dB (for five participants, one SRT with an SD greater than 3.5 dB but lower than 4 dB was included in the analyses). Three pairs of SRTs (early, late) were obtained for each noise. If the across-measures SD was higher than 3 dB and the participant was available, a fourth pair was obtained, and the across-measures SD was again calculated. If the resulting SD was still greater than 3 dB, the outlier SRT and its corresponding pair were removed from the analyses. The mean of the three or four SRTs was taken as the SRT.

The noise to be presented was chosen at random. The two temporal positions (early or late) were always administered in pairs but in random order and without removing the earphones. Participants received a break after measuring six or eight SRTs.

## Apparatus

During the measurements, participants were seated in a double-wall sound-attenuating booth and the presentation of each word was controlled by the experimenter, who was sitting outside the booth without visual interaction with the participant. Stimuli were digitally stored and presented through custom-made MATLAB (The Mathworks, version 2017a) software. Stimuli were played via an RME Fireface UCX soundcard at a sampling rate of 44.1 kHz, and with 24-bit resolution. Stimuli were presented to the listeners using ER2 insert earphones (Etymotic Research, Inc., Elk Grove Village, IL, USA), designed to give a flat frequency response at the eardrum from 250 to 8000 Hz and thus to preserve head-related spatial cues. Sound pressure levels were calibrated by placing the earphones in a Zwislocki DB-100 coupler connected to a sound-level meter. Calibration was performed at 1 kHz, and the obtained sensitivity was used at all other frequencies. The analyses of speech and noise-level fluctuations and spectra were performed in MATLAB (The Mathworks, version 2021b).

## Statistical Analyses

Statistical analyses were performed using IBM SPSS Statistics, version 23. The Shapiro–Wilk test of normality was used to examine if the SRTs followed a Gaussian distribution. We found that SRTs for natural words followed a Gaussian distribution but SRTs for vocoded words did not in some conditions. The latter result occurred because one

participant, who was tested with only vocoded words, had much more practice than the others with this type of words. Indeed, that participant's SRTs were outliers ( $>1.5 \times$  interquartile range) in 9 out of the 12 measured conditions. For this reason, the data for this participant were excluded from the analyses. The remaining SRTs followed a Gaussian distribution. Repeated measures analyses of the variance (RMANOVA) were used to test for the effect of noise type and temporal position (early vs late) on SRTs. Greenhouse–Geisser corrections were used when the sphericity assumption was violated. Bonferroni corrections were applied when performing multiple pairwise comparisons. Based on previous research (described in “Introduction”), we hypothesized that SRTs would be better (lower) for words presented late rather than early in the noise (i.e., we had a unidirectional hypothesis). Because of this, we applied one-tailed tests when assessing the effect of word temporal position (early vs late). We applied two-tailed tests for all the other comparisons.

## Results

### *Spectra and Level Fluctuations of the Noises*

We analyzed 49 noises (Table 1) to determine which ones to use to investigate the effects of noise-level fluctuations and spectrum on adaptation. Ten out of the 49 noises (indicated by an asterisk in Table 1) were excluded from the analyses because, as explained in the “Noise Classification” section, they were unlikely to mask speech. Figure 1B shows the level histograms for the remaining 39 noises when the level was calculated in nonoverlapping time windows 50 ms in duration. Each line depicts the distribution for one noise. The y axis in Figure 1C shows the proportion of levels that were within 12 dB of the peak of the distribution for each noise. According to physiological studies (Dean et al., 2005), noises with a proportion larger than 80% are expected to produce neural dynamic range adaptation to sound-level statistics. Figure 1C shows that most noises met this criterion, but some did not. Figure 1D shows the weighted spectra of the noises (gray traces) compared to that of speech (black trace with circles). Most noises had a spectrum close to that of speech and were low fluctuating ( $>80\%$  of the levels were within a 12-dB range) (bottom-left corner in Figure 1C). Therefore, most of the noises were expected to produce adaptation. A few noises, however, had a speech-like spectrum but were fluctuating (e.g., noise #47 [IFFM]) or were steady but their spectra differed from that of speech (e.g., noise #4 [Angle Grinder]). These latter noises were expected to produce little adaptation (Figure 1A). Testing adaptation for the 39 noises would be impractical for reasons of time, so we tested adaptation for six different noises chosen to be different in one dimension (temporal or spectral) but not in the other one so that we could investigate if adaptation depends on the similarity between the noise and speech spectra or on the noise-level fluctuations (see below).

### Effect of Noise Long-Term Average Spectrum on Adaptation to Noise

From all the noises analyzed (Figure 1C), we chose four of them with different long-term average spectrum but similar level fluctuations. The noises in question were shopping center (#17 in Table 1), SSN (#48), hair dryer (#1), and angle grinder (#4). These noises had 80% or more of their sound levels within a 12-dB range (Figure 2A), thus were expected to produce neural dynamic range adaptation to sound-level statistics. The spectrum of the angle grinder noise (#4), however, was more different from speech (18.6 dB) than the others. The shopping center noise or the SSN, by contrast, had the most similar spectrum to speech. If the similarity between the noise and speech spectra mattered for adaptation to noise, adaptation should be smallest for the angle grinder noise and largest for SSN or shopping center noise.

Figure 3 shows the SRTs for natural and vocoded words embedded in the four noises. Eighteen participants were tested with natural and vocoded words and three participants were tested with only natural words. Because the total number of participants tested with natural and vocoded words was different ( $N=21$  and  $18$ , respectively), we conducted separate RMANOVAs for each word type. For natural words, a two-way RMANOVA with noise type and temporal position (early and late) as factors showed a significant effect of noise type on SRT [ $F(1.5,31.5)=625.5$ ;  $p<.001$ ;  $\eta_p^2=0.976$ ]. Speech reception thresholds were similar for the SSN, shopping center noise, and hair dryer noise ( $p>.05$ ) but were significantly better (lower) for the angle grinder noise than for the other noises ( $p<.001$  for all pairwise comparisons) (Figure 3A). The RMANOVA also showed an effect of word temporal position on SRTs [ $F(1,20)=32.5$ ;  $p<.001$ ;  $\eta_p^2=0.619$ ]. Mean SRTs were better in the late than in the early condition for the SSN ( $p=.001$ ), shopping center noise ( $p=.037$ ), hair dryer noise ( $p=.007$ ), and angle grinder noise ( $p=.037$ ) (Figure 3A). The interaction between noise type and temporal position was not statistically significant [ $F(2.2,43.3)=0.3$ ;  $p=.746$ ;  $\eta_p^2=0.016$ ], indicating that adaptation to noise was not statistically different across the four noises. Mean adaptation was 0.6 dB for the SSN, 0.4 dB for the shopping center, 0.8 dB for the hair dryer noise, and 0.6 dB for the angle grinder noise (Figure 3C).

For vocoded words, a two-way RMANOVA with noise type and temporal position as factors showed a significant effect of noise type on SRT [ $F(1.7,28.6)=152.5$ ;  $p<.001$ ;  $\eta_p^2=0.900$ ]. Speech reception thresholds for the SSN were better than SRTs for the shopping center ( $p=.012$ ) or hair dryer ( $p=.004$ ) noises but were worse than SRTs for the angle grinder noise ( $p<.001$ ) (Figure 3B). Speech reception thresholds for the shopping center and hair dryer noises were not statistically different from each other ( $p=.441$ ), but they were worse than SRTs for the angle grinder noise ( $p<.001$ ). The RMANOVA also showed an effect of the temporal position on SRTs [ $F(1,17)=34.4$ ;  $p<.001$ ;  $\eta_p^2=0.669$ ].

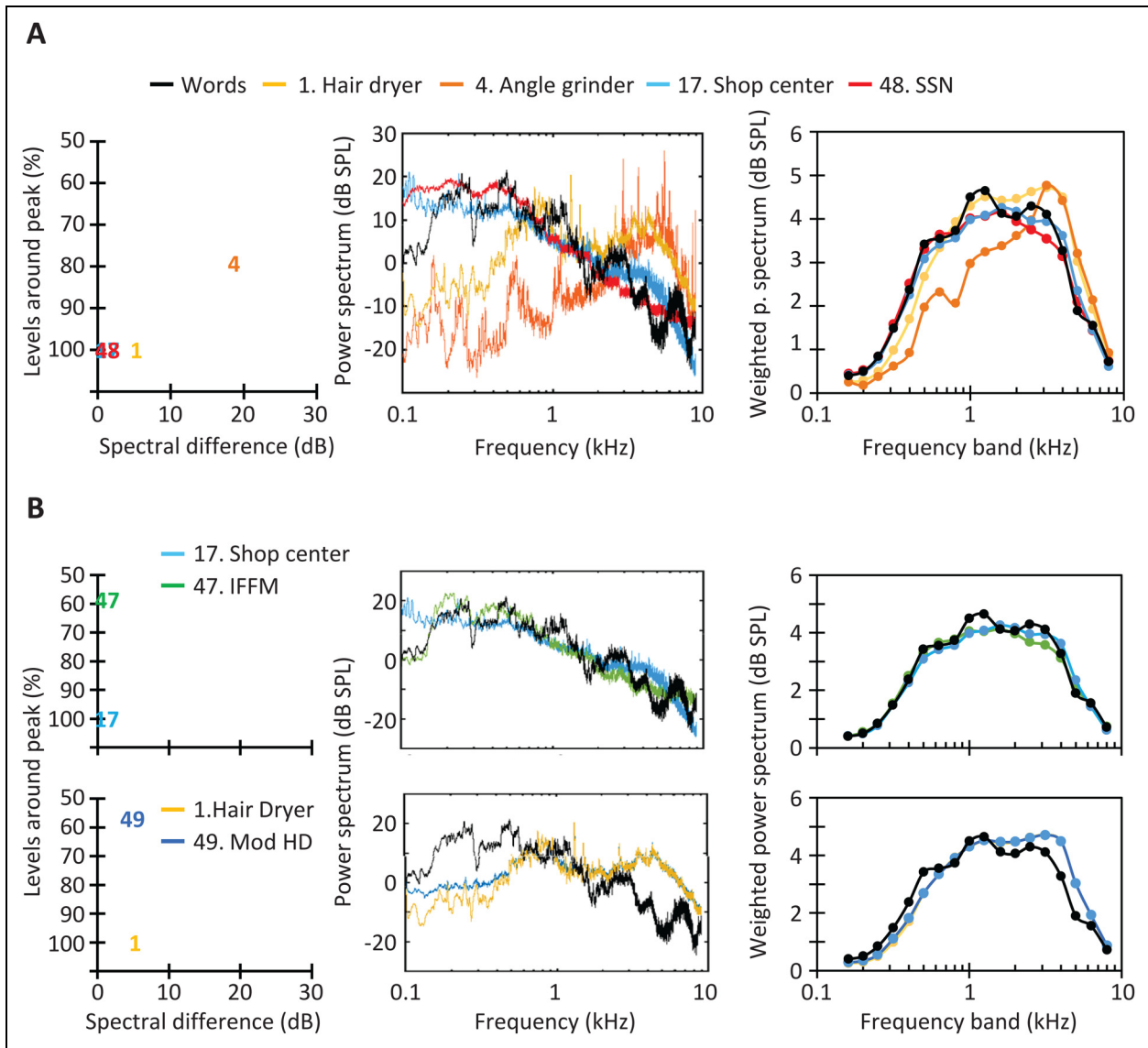
Mean SRTs were better in the late than in the early condition for the SSN ( $p<.001$ ), shopping center noise ( $p=.002$ ), hair dryer noise ( $p=.010$ ), and angle grinder noise ( $p<.001$ ) (Figure 3B). The interaction between noise type and temporal position was not statistically significant [ $F(1.9,32.6)=1.3$ ;  $p=.282$ ;  $\eta_p^2=0.071$ ]. Mean adaptation was 1.1, 1.1, 1.0, and 1.8 dB for the SSN, shopping center, hair dryer, and angle grinder noise, respectively (Figure 3C).

Altogether, these analyses confirmed that SRTs were better for the noise with the most different spectrum relative to speech (angle grinder noise), but adaptation to noise was roughly equal for all noises, thus independent from the noise spectrum.

### Effect of Noise-Level Statistics on Adaptation to Noise

To assess the effect of noise-level statistics on adaptation to noise independent from the effect of the noise long-term average spectrum, we chose to compare adaptation for two pairs of noises with similar long-term average spectra but different level fluctuations (Figure 2B). Specifically, we compared adaptation for the shopping center noise (#17) and IFFM (#47), as well as for hair dryer (#1) and modulated hair dryer noises (#49). These pairs include the least and most fluctuating noises from the noise sample (Figure 1C). If noise-level fluctuations matter for adaptation to noise, adaptation should be larger for the shopping center noise than for the IFFM because the two noises had similar speech-like spectra, but the IFFM was more fluctuating than the shopping center noise. Likewise, adaptation should be larger for the hair dryer noise than for the modulated hair dryer noise because the two noises had similar spectra (and different from speech), but the modulated hair dryer noise was more fluctuating than the hair dryer noise.

Figure 4 shows the SRTs and adaptation for these four noises. For natural words, a two-way RMANOVA with noise type (shopping center and IFFM) and temporal position (early and late) as factors was performed to test for the effects of noise type on adaptation to noise. There was a significant effect of noise type on SRT [ $F(1,20)=389.2$ ;  $p<.001$ ;  $\eta_p^2=0.951$ ]. Speech reception thresholds were better for the IFFM than for the shopping center noise (Figure 4A). The RMANOVA also showed an effect of the temporal position on SRTs [ $F(1,20)=8.3$ ;  $p=.005$ ;  $\eta_p^2=0.294$ ]. Speech reception thresholds were better in the late than in the early condition for the shopping center noise ( $p=.037$ ) and the IFFM ( $p=.035$ ). The interaction between noise type and temporal position was not statistically significant [ $F(1,20)=0.3$ ;  $p=.590$ ;  $\eta_p^2=0.015$ ], indicating that adaptation to noise was similar for the two noises. Mean adaptation was 0.4 dB for the shopping center noise and 0.7 dB for the IFFM (Figure 4C). For vocoded words (Figure 4B), the RMANOVA also revealed a significant effect of noise type [ $F(1,17)=132.3$ ;  $p<.001$ ;  $\eta_p^2=0.886$ ] and word temporal position [ $F(1,17)=24.3$ ;  $p<.001$ ;  $\eta_p^2=0.588$ ] on SRTs. Speech reception thresholds

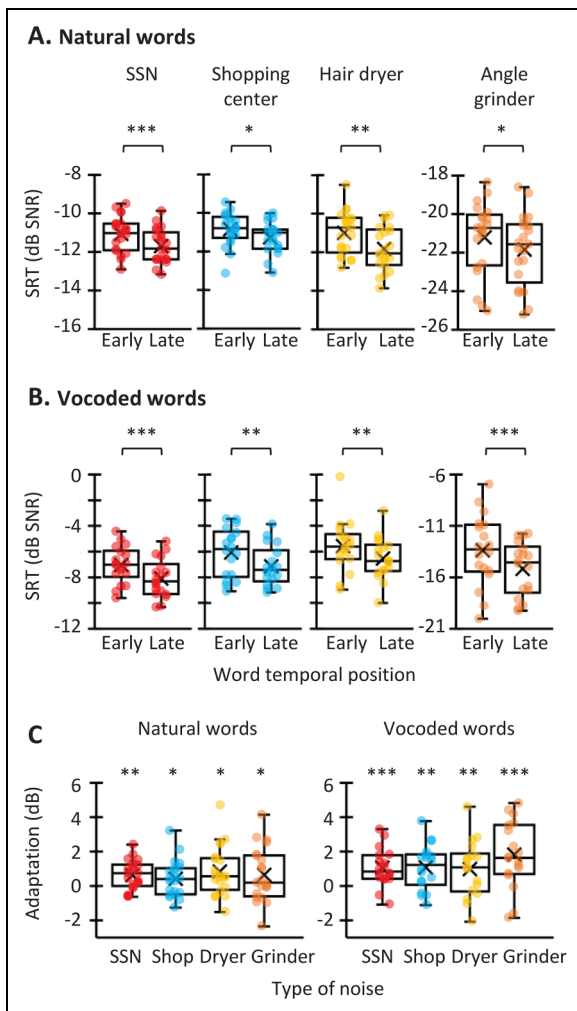


**Figure 2.** Noises Used to Measure Adaptation to Noise. (A) Noises with different spectra but close level statistics. The layout is as in Figure 1C (left panel) and Figure 1D (right panel). The mid panel shows the raw spectra (without weighting). (B) Noises with similar spectra but different level statistics.

were better for the IFFM than for the shopping center noise. Speech reception thresholds were also better in the late than in the early condition for the shopping center noise ( $p = .001$ ) and the IFFM ( $p = .003$ ). The interaction between noise type and temporal position was not statistically significant [ $F(1,18) = 0.6$ ;  $p = .439$ ;  $\eta_p^2 = 0.033$ ], indicating that adaptation to noise was not statistically different for the two noises. Mean adaptation was 1.1 dB for the shopping center noise and 1.6 dB for the IFFM (Figure 4C).

For the hair dryer noise or modulated hair dryer noise, a two-way RMANOVA showed a significant effect of noise type on SRT for natural words (Figure 4D) [ $F(1,20) = 129.3$ ;  $p < .001$ ;  $\eta_p^2 = 0.866$ ]. Speech reception thresholds were better for the modulated hair dryer than for the hair dryer

noise. The RMANOVA also showed an effect of word temporal position on SRTs [ $F(1,20) = 3.1$ ;  $p = .047$ ;  $\eta_p^2 = 0.134$ ]. Speech reception thresholds were better in the late than in the early condition for the hair dryer noise (0.8 dB;  $p = .007$ ) but not for the modulated hair dryer noise (0.0 dB;  $p = .495$ ). The interaction between temporal position and noise type was not statistically significant [ $F(1,20) = 4.0$ ;  $p = .058$ ;  $\eta_p^2 = 0.168$ ]. For vocoded words (Figure 4E), the RMANOVA showed a significant effect of noise type [ $F(1,17) = 16.5$ ;  $p = .001$ ;  $\eta_p^2 = 0.492$ ] and word temporal position [ $F(1,17) = 28.9$ ;  $p < .001$ ;  $\eta_p^2 = 0.629$ ] on SRTs. Speech reception thresholds were better for the modulated hair dryer noise than for the hair dryer noise. Speech reception thresholds were also better in the late than in the early



**Figure 3.** Effect of Noise Long-Term Average spectrum on Adaptation to Noise. (A) Speech reception thresholds (SRTs) for natural words ( $N=21$ ) embedded in four different noises and presented in the early and late conditions. Bottom, middle, and top lines in each box plot indicate the 25th, 50th (median), and 75th percentiles, respectively. Crosses represent mean values. Dots indicate individual results. (B) As panel A but for vocoded words ( $N=18$ ). (C) Adaptation to noise for natural and vocoded words. Adaptation was calculated as the difference between SRTs in the early and late conditions. Positive values indicate better (lower) SRTs in the late than in the early condition. \* $p \leq .05$ ; \*\* $p \leq .01$ ; \*\*\* $p \leq .001$ .

condition for the hair dryer ( $p = .010$ ) and modulated hair dryer noises ( $p = .001$ ). The interaction between noise type and temporal position was not statistically significant [ $F(1,17) = 0.5$ ;  $p = .513$ ;  $\eta_p^2 = 0.026$ ]. Mean adaptation was 1.0 dB for the hair dryer noise and 1.4 dB for the modulated hair dryer noise (Figure 4F).

### Adaptation Across All Noises

To summarize the results and to test our hypothesis (illustrated in Figure 1A), we pooled together adaptation for the

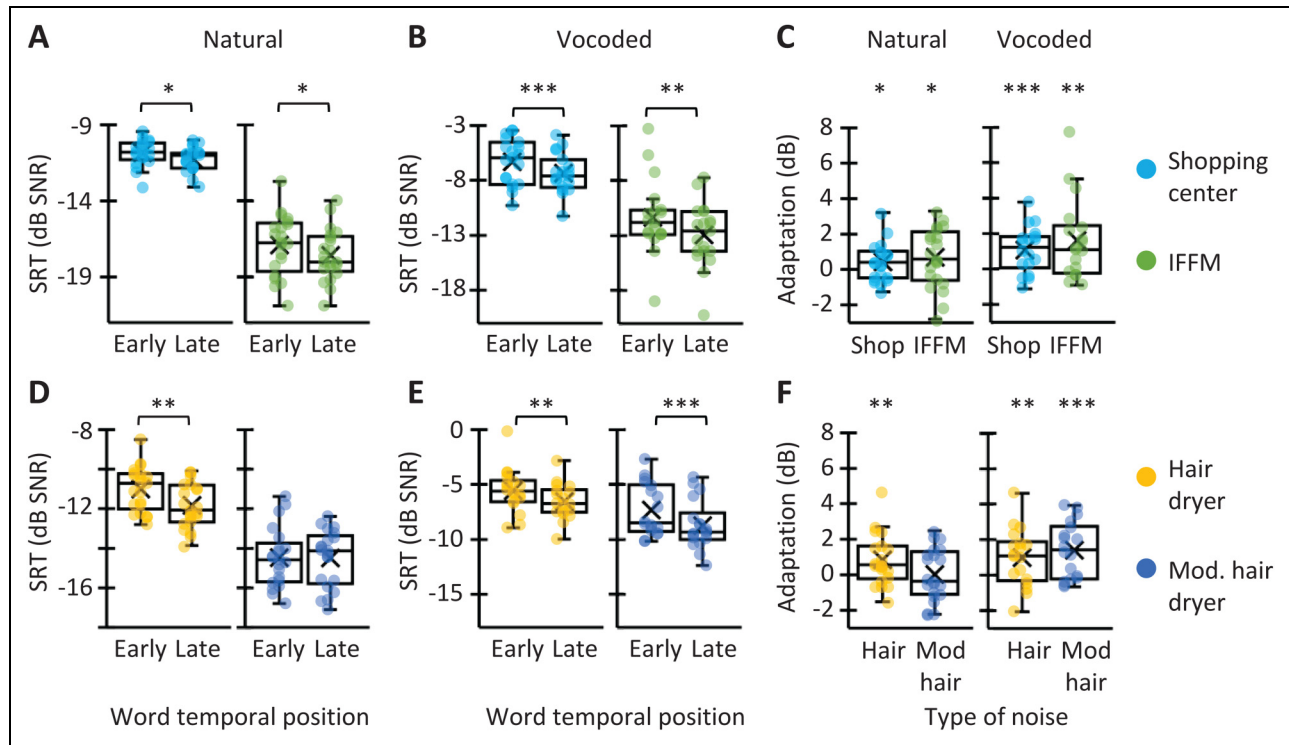
six noises and separately for natural and vocoded words (Figure 5). A two-way RMANOVA with word type and noise type as factors confirmed that adaptation was larger for vocoded words than for natural words [ $F(1,17) = 10.5$ ;  $p = .005$ ;  $\eta_p^2 = 0.380$ ], consistent with our previous studies (Marrufo-Pérez et al., 2018, 2020). The mean (across-noise) adaptation was 0.5 for natural words ( $N=21$ ) and 1.3 dB for vocoded words ( $N=18$ ). Adaptation was not statistically different across the six noises [ $F(5,85) = 0.6$ ;  $p = .666$ ;  $\eta_p^2 = 0.037$ ], and there was no interaction between word type and noise type [ $F(5,85) = 0.9$ ;  $p = .480$ ;  $\eta_p^2 = 0.051$ ]. Therefore, contrary to our hypothesis, adaptation was independent of the noise-level fluctuations and of noise-to-speech spectral difference.

### Discussion

We investigated if and to what extent adaptation to noise in speech recognition occurs for realistic and synthetic noises. Noises were classified in terms of their long-term average spectrum and their level fluctuations. Recognition and adaptation were assessed for six noises that could potentially mask speech while covering a wide range of level fluctuations and spectral distance from speech (Figure 1C, Figure 2). The best SRTs were found for the angle grinder noise, whose spectrum was the most different to speech (Figure 2A). This is consistent with previous studies (Fletcher, 1940; Patterson, 1976). In addition, SRTs were better when speech was embedded in fluctuating than in steady noise (Figure 4), which is consistent with the phenomenon of masking release (Gnansia et al., 2008, 2009; Lorenzi et al., 2006; Miller & Licklider, 1950). Adaptation to noise, however, was independent from the long-term average spectrum (Figures 3 and 5) or level distribution (Figures 4 and 5) of the noise. These findings do not support the hypotheses that adaptation would be larger for steady than for fluctuating noises, and for noises whose spectra are closer to the speech spectrum.

### On the Similar Adaptation Across the Different Noise Backgrounds

Previous studies have found an effect of noise long-term average spectrum on adaptation to noise (Cervera & Ainsworth, 2005; Cervera & Gonzalez-Alvarez, 2007; Khalighinejad et al., 2019). Cervera and Gonzalez-Alvarez (2007) presented bandpass-filtered syllables (0.92–2 kHz) and found more adaptation when the noise was filtered with the same bandpass filter as speech than when it was filtered with a narrower or a broader bandpass filter. Khalighinejad et al. (2019) found behavioral adaptation in phoneme recognition for the “city” and “bar” noise but not for the “jet” noise. The long-term average spectrum of the jet noise was the most different from speech. Here, we did not find an effect of noise spectrum on adaptation to noise (Figure 5). There are at least two differences

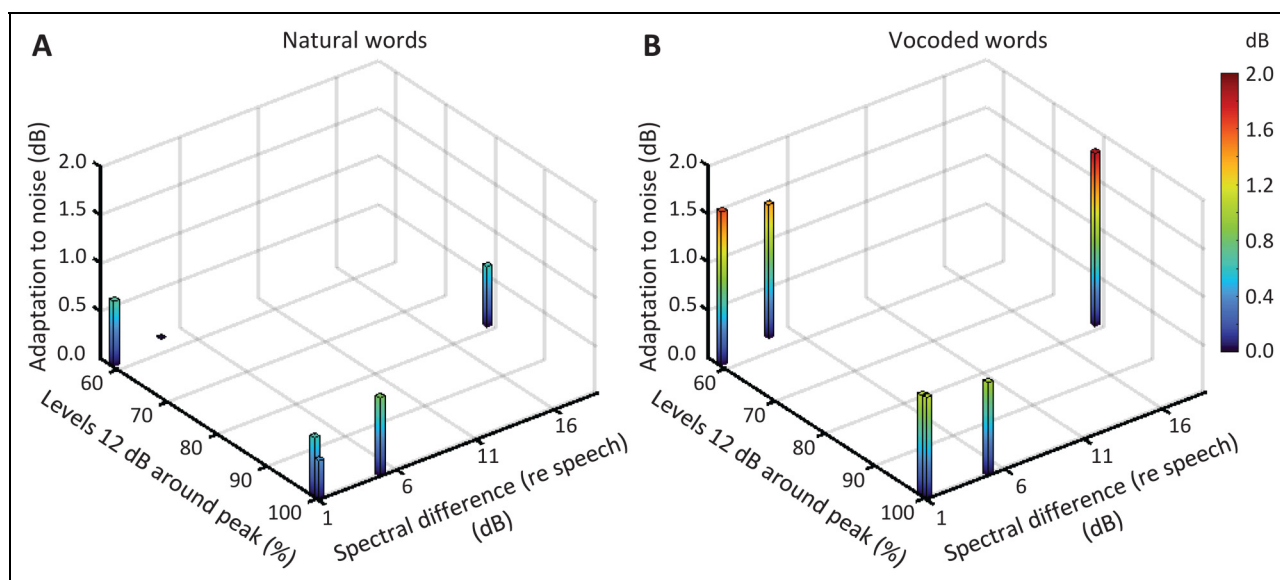


**Figure 4.** Effect of Noise-Level Fluctuations on Adaptation to Noise. (A) Speech reception thresholds for natural words embedded in the shopping center noise or International Female Fluctuating Masker (IFFM) and presented early or late in the noise. (B) The same as panel A but for vocoded words. (C) Adaptation to noise for natural and vocoded words for the shopping center noise and IFFM. (D) Speech reception thresholds for natural words embedded in the hair dryer noise and modulated hair dryer noise. (E) As D but for vocoded words. (F) Adaptation to noise for natural and vocoded words for the hair dryer and modulated hair dryer noises. Bottom, middle, and top lines in each box plot indicate the 25th, 50th (median), and 75th percentiles, respectively. Crosses represent mean values. Dots indicate individual results. \* $p \leq .05$ ; \*\* $p \leq .01$ ; \*\*\* $p \leq .001$ .

between the cited studies and the present one. First, previous studies focused on investigating syllable recognition instead of word recognition. It might be that noise adaptation differences due to noise spectrum only occur in phoneme recognition. Second, and more importantly, the cited studies presented speech at a fixed SNR and measured the percentage of recognized syllables presented early and late in the noise. Here, we varied the SNR to find 50% word recognition for all noises. Our hypothesis was that adaptation of the dynamic range of auditory neurons toward the level of the noise preceding the word (Dean et al., 2005, 2008; Watkins & Barbour, 2008; Wen et al., 2009, 2012) would facilitate the encoding of speech envelope and spectral cues (Ainsworth & Meyer, 1994; Marrufo-Pérez & Lopez-Poveda, 2022). If this was the actual mechanism underlying behavioral adaptation, then behavioral adaptation would depend on the overlap between the neural excitation patterns produced by the noise and speech. Therefore, for a fixed SNR, one would expect less adaptation to noise by increasing the noise-to-speech spectral difference because of the lesser overlap between the excitation patterns of the noise and speech. However, if all noises are forced to produce the same amount of masking (i.e., 50% speech recognition), the overlap between the neural excitation

patterns produced by the noise and the speech should be constant for all noises, thus resulting in similar adaptation across noises. Further research is needed to investigate this possibility.

Our second hypothesis was that, because neural dynamic range adaptation toward the noise level occurs when the noise level is steady (or change so slowly over time that neurons adapt to the statistical changes; Willmore et al., 2014), adaptation to noise would occur for steady but not for fluctuating noises. We, however, did not find differences in adaptation across noises with markedly different fluctuations (Figure 5). This finding appears to be inconsistent with an earlier study where we found adaptation when the noise preceding the word was steady but not fluctuating in level (Marrufo-Pérez et al., 2020). One difference between the two studies is that in our previous study we artificially modulated the precursor noise by changing its level every 50 ms so that the modulation was close to the modulations used in physiological studies (e.g., in Dean et al., 2005). Here, by contrast, the modulated noises (IFFM and modulated hair dryer noise) had the natural modulations present in speech, with a peak in modulation spectrum around 5 Hz (Ding et al., 2017). It might be that these low modulations of the



**Figure 5.** Mean Adaptation for the Six Noises Used in the Study. Data are for natural (A) and vocoded words (B) (replotted from Figures 3 and 4). Note that there are six columns (one per noise) in each panel, but two columns are next to each other at the bottom corner of the graph.

masker are slow enough to allow neurons to adapt to the statistical changes, in contrast to the faster changes that occur when the level changes every 50 ms (modulation rate = 20 Hz). In addition, in our previous study, we modified the level fluctuations of the precursor noise while the noise simultaneous to the word was steady. Here, by contrast, the precursor and simultaneous noise were the same. Khalighinejad et al. (2019) found that speech is worse represented in the auditory cortex just after changing the type of background noise than when speech is presented in the same type of continuous noise for a while, which may explain why we did not find adaptation in our previous study (Marrufo-Pérez et al., 2020) but have found it here. Moreover, Khalighinejad et al. (2019) found that after switching on new background noise, the auditory cortex adapts to enhance the response to speech with a time course of 420 ms. Importantly, they found that this adaptation occurred for the three noises they used regardless of their level fluctuations or spectral profiles, which is consistent with our results.

Our study was designed around the phenomenon of noise adaptation, so the results have been interpreted accordingly. However, our results might also be regarded as consistent with the phenomenon of stream segregation (McDermott et al., 2011; Sohoglu & Chait, 2016). For instance, Sohoglu and Chait (2016) showed that the response of cortical auditory regions to tones is larger when tones are presented within predictable (regular) acoustic scenes than within unpredictable acoustic scenes, even when listeners did not attend to the stimuli. The results were interpreted as evidence that the auditory system models ongoing scenes and that novel events that violate these models are perceived as salient. If the auditory system needs time to create a stable representation of the noise

based on its ongoing statistics, then the speech and noise would be more easily segregated from each other in the late than in the early condition, which would result in behavioral noise adaptation. According to this interpretation, the noise spectrum, or its level statistics per se would not matter; what would matter is how predictable the different noises are. If some or all the characteristics of the noises used here were stable enough that the listener can create a model for the ongoing noise (Hicks & McDermott, 2024), then adaptation should occur for all noises, although it is yet to be demonstrated if this is the case.

### *On the Larger Adaptation for Vocoded Than for Natural Words*

Adaptation to noise was larger for vocoded words than for natural words (Figure 5). This result has been consistently reported in previous studies (e.g., López-Ramos et al., 2024; Marrufo-Pérez et al., 2018, 2020) but the reason for it is uncertain. Marrufo-Pérez et al. (2020) conjectured that it may be because adaptation enhances the speech envelope and spectral cues, but not the TFS cues. Therefore, when presented with natural words, listeners would show small adaptation because they can recognize natural words using a cue (TFS) that is not enhanced by adaptation. For tone-vocoded words, by contrast, TFS cues are scarce (or absent), thus the enhancement of envelope and spectral cues has a larger effect in the recognition of these words. This hypothesis seems to be supported by the study of López-Ramos et al. (2024), who found adaptation in spectral and temporal modulation detection but not in spectrotemporal modulation detection

presumably because TFS plays a role in the latter. Either way, the different adaptation for natural and vocoded speech suggests that the creation of a model (schema) of the ongoing noise (Hicks & McDermott, 2024) may not be the only mechanism behind the improvement in recognition when words are delayed in the noise. That is, because in the present study the background noise was always natural, listeners should have formed equal noise schemas when they were presented with natural and vocoded speech. Therefore, if adaptation resulted solely from noise schemas, adaptation should have been similar for the two types of words, and this is not the case.

### On the Importance of Adaptation to Noise

Across the six tested noises, mean adaptation was 0.5 dB for natural words and 1.3 dB for vocoded words (Figure 5). One might argue that these values are small for adaptation to play a significant role in everyday listening situations. However, several considerations are in order. First, a 1–2 dB improvement in SRT in noise corresponds to about 15–25% improvement in syllable or word recognition at a fixed SNR (Ben-David et al., 2016; Cervera & Gonzalez-Alvarez, 2007), which is arguably meaningful. Second, adaptation varies across listeners. For some listeners, adaptation can be as large as 3–4 dB (Figures 3C, 4C and F), and some of them show substantial adaptation for all noises (Supplemental Figure S2). This suggests that adaptation can improve speech recognition for some listeners. The reason for the variability in adaptation across listeners is yet to be explained (see Marrufo-Pérez and Lopez-Poveda, 2022, for a review). Lastly, the average duration of the disyllabic words used in the present study was ~650 ms. Because neural adaptation to sound-level statistics occurs with a time course of ~400 ms (Wen et al., 2012), significant neural adaptation could have occurred toward the end of each word in the early condition. This may have caused adaptation to be smaller than expected if shorter speech tokens, such as syllables or phonemes, had been used. In summary, the present study shows that meaningful adaptation can occur for realistic noises when the SNR is appropriate (i.e., different across noises). This opens the possibility for adaptation to occur in realistic situations, such as when there is continuous background noise and speech starts after a pause. It is expected, however, that some listeners benefit more from adaptation in such situations than others.

### Limitations and Implications

The tested noises, which had different spectra and level fluctuations, produced different SRTs (up to 11 dB difference on average) but not different adaptation to noise. It remains uncertain, however, whether adaptation would occur for noises with a spectrum more different from speech than those used here. Also, it remains uncertain whether adaptation would be noise dependent when the SNR is fixed across noises,

although the behavioral results from Khalighinejad et al. (2019) suggest that this might be the case.

A possible limitation of the current study relates to the word-onset delay. Ben-David et al. (2012) demonstrated that intelligibility does not improve further when words are delayed beyond ~600 ms in the noise. Accordingly, we delayed the words by 800 ms to ensure maximum adaptation. Results from Hicks and McDermott (2024), however, suggest that a 912-ms delay is necessary to reach the plateau in the improvement, at least when the task involves detecting or recognizing non-speech sounds. It is uncertain whether adaptation may have been greater for the noises used in our study if the delay had been longer than 800 ms.

## Conclusions

1. Normal-hearing listeners can adapt to realistic and nonrealistic noises, that is, their SRTs for disyllabic words in noise improve when the words are delayed 800 ms in the noise.
2. Adaptation to noise in a word recognition task is independent from the long-term average spectrum of the noise or its level fluctuations.
3. Average (across noises) adaptation was 0.5 and 1.3 dB for natural and vocoded words, respectively, with some listeners showing improvements of up to ~4 dB.

## Abbreviations

HL	hearing level
HRTF	head-related transfer functions
IFFM	International Female Fluctuating Masker
NH	normal hearing
RMANOVA	repeated measures analyses of the variance
SII	Speech Intelligibility Index
SNR	signal-to-noise ratio
SRT	speech reception threshold
SSN	speech-shaped noise
TFS	temporal fine structure

## Acknowledgments

A portion of the data was presented as the master thesis of Brigitte Escobar Campuzano (2021, Universidad de Salamanca). The authors thank her and Milagros J. Fumero for help with data collection.

## Declaration of Conflicting Interests



The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This

work was supported by the European Regional Development Fund and Ministerio de Ciencia e Innovación (grant number PID2019-108985GB-I00).

## ORCID iDs

Miriam I. Marrufo-Pérez  <https://orcid.org/0000-0002-8737-8107>  
 Enrique A. Lopez-Poveda  <https://orcid.org/0000-0002-6886-154X>

## Supplemental Material

Supplemental material for this article is available online.

## References

- Acoustical Society of America (ASA), ANSI/ASA S3.5-1997. (R2007). *Methods for calculation of the speech intelligibility index*. Acoustical Society of America.
- Ainsworth, W. A., & Meyer, G. F. (1994). Recognition of plosive syllables in noise: Comparison of an auditory model with human performance. *Journal of the Acoustical Society of America*, 96, 687–694. <https://doi.org/10.1121/1.410306>
- American National Standards Institute. (1996). *S3.6 Specification for audiometers*. American National Standards Institute.
- Ben-David, B. M., Avivi-Reich, M., & Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hearing Research*, 341, 9–18. <https://doi.org/10.1016/j.heares.2016.07.016>
- Ben-David, B. M., Tse, V. Y., & Schneider, B. A. (2012). Does it take older adults longer than younger adults to perceptually segregate a speech target from a background masker? *Hearing Research*, 290, 55–63. <https://doi.org/10.1016/j.heares.2012.04.022>
- Bjerg, A. P., & Larsen, J. N. (2006). *Recording of natural sounds for hearing aid measurements and fitting acoustic technology*, Ørsted. Technical University of Denmark, 245.
- Cárdenas, M. R., & Marrero, V. (1994). *Cuaderno de logaudiometría*. Universidad Nacional de Educación a Distancia.
- Cervera, T., & Ainsworth, W. A. (2005). Effects of preceding noise on the perception of voiced plosives. *Acta Acustica United With Acustica*, 91, 132–144.
- Cervera, T., & Gonzalez-Alvarez, J. (2007). Temporal effects of preceding bandpass and band-stop noise on the recognition of voiced stops. *Acta Acustica United With Acustica*, 93, 1036–1045.
- Dean, I., Harper, N. S., & McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nature Neuroscience*, 8, 1684–1689. <https://doi.org/10.1038/nn1541>
- Dean, I., Robinson, B. L., Harper, N. S., & McAlpine, D. (2008). Rapid neural adaptation to sound level statistics. *Journal of Neuroscience*, 28, 6430–6438. <https://doi.org/10.1523/JNEUROSCI.0470-08.2008>
- Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*, 81(Pt B), 181–187. <https://doi.org/10.1016/j.neubiorev.2017.02.011>
- Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47–61. <https://doi.org/10.1103/RevModPhys.12.47>
- Gardner, W. G., & Martin, K. D. (1995). HRTF measurements of a KEMAR. *Journal of the Acoustical Society of America*, 97, 3907–3908. <https://doi.org/10.1121/1.412407>
- Gnansia, D., Jourdes, V., & Lorenzi, C. (2008). Effect of masker modulation depth on speech masking release. *Hearing Research*, 239, 60–68. <https://doi.org/10.1016/j.heares.2008.01.012>
- Gnansia, D., Péan, V., Meyer, B., & Lorenzi, C. (2009). Effects of spectral smearing and temporal fine structure degradation on speech masking release. *Journal of the Acoustical Society of America*, 125(6), 4023–4033. <https://doi.org/10.1121/1.3126344>
- Hicks, J. A., & McDermott, J. H. (2024). Noise schemas aid hearing in noise. *Proceedings of the National Academy of Sciences*, 121(47), e2408995121. <https://doi.org/10.1073/pnas.2408995121>
- Holube, I. (2011). Speech intelligibility in fluctuating maskers. In International Symposium on Auditory and Audiological Research (ISAAR), Nyborg, Denmark.
- Khalighinejad, B., Herrero, J. L., Mehta, A. D., & Mesgarani, N. (2019). Adaptation of the human auditory cortex to changing background noise. *Nature Communications*, 10(1), 2509. <https://doi.org/10.1038/s41467-019-10611-4>
- King, A. J., & Walker, K. M. (2020). Listening in complex acoustic scenes. *Current Opinion in Physiology*, 18, 63–72. <https://doi.org/10.1016/j.cophys.2020.09.001>
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–677. <https://doi.org/10.1121/1.1912375>
- López-Ramos, D., Marrufo-Pérez, M. I., Eustaquio-Martín, A., López-Bascuas, L. E., & Lopez-Poveda, E. A. (2024). Adaptation to noise in spectrotemporal modulation detection and word recognition. *Trends in Hearing*, 28, 23312165241266322. <https://doi.org/10.1177/23312165241266322>
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103, 18866–18869. <https://doi.org/10.1073/pnas.0607364103>
- Marrufo-Pérez, M. I., Eustaquio-Martín, A., & Lopez-Poveda, E. A. (2018). Adaptation to noise in human speech recognition unrelated to the medial olivocochlear reflex. *Journal of Neuroscience*, 38(17), 4138–4145. <https://doi.org/10.1523/JNEUROSCI.0024-18.2018>
- Marrufo-Pérez, M. I., & Lopez-Poveda, E. A. (2022). Adaptation to noise in normal and impaired hearing. *Journal of the Acoustical Society of America*, 151(3), 1741–1753. <https://doi.org/10.1121/10.0009802>
- Marrufo-Pérez, M. I., Sturla-Carreto, D. P., Eustaquio-Martín, A., & Lopez-Poveda, E. A. (2020). Adaptation to noise in human speech recognition depends on noise-level statistics and fast dynamic-range compression. *Journal of Neuroscience*, 40(34), 6613–6623. <https://doi.org/10.1523/JNEUROSCI.0469-20.2020>

- McDermott, J. H., Wroblewski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences*, 108, 1188–1193. <https://doi.org/10.1073/pnas.1004765108>
- Miller, G. A., & Licklider, J. (1950). The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*, 22, 167–173. <https://doi.org/10.1121/1.1906584>
- Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95, 1085–1099. <https://doi.org/10.1121/1.408469>
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59(3), 640–654. <https://doi.org/10.1121/1.380914>
- Rabinowitz, N. C., Willmore, B. D., King, A. J., & Schnupp, J. W. (2013). Constructing noise-invariant representations of sound in the auditory pathway. *PLoS Biology*, 11(11), e1001710. <https://doi.org/10.1371/journal.pbio.1001710>
- Snidero, T., Jacobsen, F., & Buchholz, J. (2011). *Measuring HRTFs of Brüel & Kjær Type 4128-C, G.R.A.S. KEMAR Type 45BM, and Head Acoustics HMS II.3 Head and Torso Simulators*. Technical University of Denmark, Department of Electrical Engineering.
- Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *Elife*, 5, 1–17. <https://doi.org/10.7554/eLife.19113>
- Watkins, P. V., & Barbour, D. L. (2008). Specialized neuronal adaptation for preserving input sensitivity. *Nature Neuroscience*, 11, 1259–1261. <https://doi.org/10.1038/nn.2201>
- Wen, B., Wang, G. I., Dean, I., & Delgutte, B. (2009). Dynamic range adaptation to sound level statistics in the auditory nerve. *Journal of Neuroscience*, 29, 13797–13808. <https://doi.org/10.1523/JNEUROSCI.5610-08.2009>
- Wen, B., Wang, G. I., Dean, I., & Delgutte, B. (2012). Time course of dynamic range adaptation in the auditory nerve. *Journal of Neurophysiology*, 108, 69–82. <https://doi.org/10.1152/jn.00055.2012>
- Willmore, B. D. B., Cooke, J. E., & King, A. J. (2014). Hearing in noisy environments: Noise invariance and contrast gain control. *Journal of Physiology*, 592, 3371–3381. <https://doi.org/10.1113/jphysiol.2014.274886>