Research Article

JARO
Journal of the Association for Research in Otolaryngology

Effects of Noise Reduction on AM and FM Perception

D. Timothy Ives,[1] Axelle Calcus,[1] Sridhar Kalluri,[2] Olaf Strelcyk,[2] Stanley Sheft,[3] and Christian Lorenzi[1]

[1]*Ecole normale supérieure, Institut d'Etude de la Cognition, Paris Sciences et Lettres, 75005 Paris, France*

[2]*Starkey Hearing Research Center, Berkeley, CA 94704, USA*

[3]*Communication Disorders and Sciences, Rush University Medical Center, Chicago, IL 60612, USA*

ABSTRACT

The goal of noise reduction (NR) algorithms in digital hearing aid devices is to reduce background noise whilst preserving as much of the original signal as possible. These algorithms may increase the signal-to-noise ratio (SNR) in an ideal case, but they generally fail to improve speech intelligibility. However, due to the complex nature of speech, it is difficult to disentangle the numerous low- and high-level effects of NR that may underlie the lack of speech perception benefits. The goal of this study was to better understand why NR algorithms do not improve speech intelligibility by investigating the effects of NR on the ability to discriminate two basic acoustic features, namely amplitude modulation (AM) and frequency modulation (FM) cues, known to be crucial for speech identification in quiet and in noise. Here, discrimination of complex, non-linguistic AM and FM patterns was measured for normal hearing listeners using a same/different task. The stimuli were generated by modulating 1-kHz pure tones by either a two-component AM or FM modulator with patterns changed by manipulating component phases. Modulation rates were centered on 3 Hz. Discrimination of AM and FM patterns was measured in quiet and in the presence of a white noise that had been passed through a gammatone filter centered on 1 kHz. The noise was presented at SNRs ranging from −6 to +12 dB. Stimuli were left as such or processed via an NR algorithm based on the spectral subtraction method. NR was found to yield small but systematic improvements in discrimination for the AM conditions at favorable SNRs but had little effect, if any, on FM discrimination. A computational model of early auditory processing was developed to quantify the fidelity of AM and FM transmission. The model captured the improvement in discrimination performance for AM stimuli at high SNRs with NR. However, the model also predicted a relatively small detrimental effect of NR for FM stimuli in contrast with the average psychophysical data. Overall, these results suggest that the lack of benefits of NR on speech intelligibility is partly caused by the limited effect of NR on the transmission of narrowband speech modulation cues.

Keywords: noise reduction, spectral subtraction, amplitude modulation, frequency modulation, pattern discrimination

INTRODUCTION

Noise reduction (NR) algorithms employed in modern digital hearing aids fall short of their goal: the processing of speech in the presence of background noise by such algorithms does not improve comprehension compared to the unprocessed conditions. Hu and Loizou ([2007](#)) investigated the effectiveness of eight NR algorithms using a large set of 40 normal hearing (NH) listeners. They measured sentence intelligibility scores for each of the eight NR algorithms using four types of noise maskers presented at two signal-to-noise ratios (SNRs). Their findings showed that in all but one of the 64 conditions (i.e., one algorithm with one type of noise at one SNR), NR algorithms did not improve

Correspondence to: D. Timothy Ives · Ecole normale supérieure · Institut d'Etude de la Cognition · Paris Sciences et Lettres, 75005 Paris, France. Telephone: +3314432267; fax: +33 1 44 32 20 99 email: Tim.Ives@ens.fr

sentence recognition. Furthermore, Sarampalis et al. (2009) showed that NR algorithms based on spectral subtraction, representative of current technology (Ephraim and Malah 1984, 1985), even slightly reduced word recognition with high-context sentences.

Spectral subtractive algorithms are based on an assumption of additive noise; they estimate the original spectrum of the speech signal by subtracting an estimate of the noise spectrum from the noisy speech spectrum. Dubbelboer and Houtgast (2007) showed that spectral subtractive NR algorithms may improve the transmission of the original narrowband speech amplitude modulation (AM) cues in noise by increasing their strength. Several factors may limit the ability of listeners to make use of these modulation cues following NR processing of the signal. As demonstrated by Dubbelboer and Houtgast (2007), NR does not reduce the detrimental effect of either stochastic AM fluctuations of noise or of the coincidental phase interactions between speech and noise (that is, the corruption of speech frequency modulation (FM) cues). Moreover, the subtraction algorithms can make a bad estimate of the background noise due to either (1) the use of a single microphone system in which the noise estimate must be separated from a combined signal plus noise mixture or (2) the use of a spectral representation of the noise that is too coarse to capture relevant features. If the noise is poorly estimated, negative values can be produced in the signal envelope when a constant noise level is subtracted from the signal envelope (i.e., when the signal envelope is lower than the long-term noise level). Subsequent processing of the envelope sets these negative values to zero. This in turn produces distortions in the audio and modulation domains, both of which carry important information for speech recognition in quiet and in noise (e.g., French and Steinberg 1947; Houtgast and Steeneken 1985; Shannon et al. 1995; Zeng et al. 2005).

Speech is commonly used to assess the effectiveness of NR algorithms (e.g., Hu and Loizou 2007; Dubbelboer and Houtgast 2007; Jørgensen and Dau 2011); however, speech is a complex stimulus for which recognition relies on the operation of a number of low- and high-level auditory and/or cognitive processes (Moore 2008; Darwin 2008). Consequently, it is difficult to disentangle the effects of the NR algorithms on these numerous processes. Also, speech is over-learned and as such one would need to separate the contribution of linguistic effects from acoustical changes in the signal in determining the basis of any improvement or deficit in performance.

In this study, we measured the effectiveness of an NR algorithm based on the spectral subtraction method, which uses a single microphone input. To avoid the confounding cognitive effects that may arise when using speech stimuli, we used well-controlled, non-linguistic stimuli based on Ives and Lorenzi (2011) (see also Ardoint et al. 2008 for a related paradigm). These stimuli exhibited complex modulation patterns of either amplitude or frequency, with temporal characteristics comparable to those of speech (Houtgast and Steeneken 1985; Sheft and Lorenzi 2008). Our stimuli allowed us to assess whether or not the NR algorithms significantly disrupt the transmission of narrowband AM and FM information relevant to speech recognition in quiet and in noise (cf. Shannon et al. 1995; Zeng et al. 2005), and thus address the concerns raised by Dubbelboer and Houtgast (2007), by assessing the effect of NR on the discrimination of modulation patterns.

METHOD

Participants

Nine NH listeners participated. All of the listeners showed audiometric thresholds not exceeding 20 dB hearing level at frequencies of 0.25, 0.5, 1, 2, and 4 kHz. The age range was 22–32 years with a mean of 25.3 years. All stimuli were delivered monaurally to the right ear at 65 dB SPL via Sennheiser HD250 Linear II headphones. Listeners were seated in a double-walled IAC sound isolation booth.

Stimuli and modulation discrimination tasks

The tasks required listeners to discriminate pairs of stimuli that differed in their modulation patterns. All stimuli used in the present modulation discrimination tasks were 1-kHz pure tone carriers (with random starting phase) modulated either in amplitude (AM condition) or frequency (FM condition) by a complex modulator consisting of two sinusoids. Equations 1 and 2 describe the AM and FM modulators, respectively:

$$S_{AM}(t) = 1 + m\sin(2\pi f_1 t + \theta_1) + m\,\sin(2\pi f_2 t + \theta_2) \quad (1)$$

$$S_{FM}(t) = \beta\,\sin(2\pi f_1 t + \theta_1) + \beta\,\sin(2\pi f_2 t + \theta_2), \quad (2)$$

where t corresponds to time (in seconds). The frequencies of these two sinusoids, f_1 and f_2, were inharmonically related ($f_2 = 1.254 \times f_1$) and centered symmetrically on a logarithmic axis about a nominal center frequency (f_c) of 3 Hz, chosen because it corresponds to the most salient and critical modulation rate in the production and understanding of continuous speech (Steeneken and Houtgast 1980; Houtgast and Steeneken 1985; see also Sheft et al.

(2012) for FM cues in speech). In order to prevent listeners from building templates of the stimuli, fc was roved by ±0.25 octave (i.e., 0.5 octave range) across trials. Within a trial, fc was constant for the stimuli of both intervals. The starting phases of the two sinusoids (θ_1, θ_2) of the modulators were taken from two sets of 136 randomly generated phase pairs. On each trial, θ_1 and θ_2 were either identical or different across the two intervals, i.e., $\theta_{1\ \text{interval 1}} = \theta_{1\ \text{interval 2}}$ and $\theta_{2\ \text{interval 1}} = \theta_{2\ \text{interval 2}}$ for the identical trials and $\theta_{1\ \text{interval 1}} \diamond \theta_{1\ \text{interval 2}}$ and $\theta_{2\ \text{interval 1}} \diamond \theta_{2\ \text{interval 2}}$ for the different trials, thereby producing either *identical* or *different* pairs of complex modulation patterns, respectively. The duration of each complex modulation pattern was set to one pseudo period of the complex modulator [stimulus duration$=1/(f_2-f_1)$]. As a consequence, the duration of the stimuli varied across trials (but was identical within trials). Typically, the duration of each stimulus within the pair was between 1.25 and 1.75 s with a mean value of 1.5 s. Stimuli were ramped on and off using a raised half-period cosine function with a duration of 100 ms. The interstimulus interval was 500 ms. For the AM task, the magnitude (and thus, modulation depth) of each sinusoid, m, was fixed at 0.5. For the FM task, the value of the modulation index, β (peak frequency deviation/ maximum modulation frequency), for each sinusoid was set to 5. Together with the frequency roving, this resulted in the instantaneous frequency varying between 960 and 1,040 Hz. All stimuli were equated in root mean square (rms) power before presentation to the listeners.

Discrimination of AM and FM patterns was measured in quiet and in the presence of a simultaneous white noise masker that had been passed through a gammatone filter (Patterson et al. 1987) centered on 1 kHz. The noise was presented at SNRs ranging from −6 to +12 dB (SNR being calculated using the rms of the narrowband signal and noise). For each stimulus interval within a trial, a different noise token was used.

Discrimination performance in terms of d′ scores for the AM and FM conditions was measured by means of a same–different procedure (see Macmillan and Creelman 2005). Discrimination, d′ scores, were calculated by taking the difference between the Z-scores of the hit rates (HR) and the Z-scores of the false alarm rates (FAR). Extreme values for either the HR or the FAR were corrected as specified by Macmillan and Creelman (2005), i.e., half a hit or half a false alarm was added (if HR or FAR was 0) or subtracted (if HR or FAR was 1). Feedback as to the correct answer was provided to the listener at the end of each trial. Discrimination performance was measured in separate blocks for each type of modulation (AM, FM) and for each experimental condition. Fourteen conditions were tested with nine listeners and a further four conditions (those using an SNR of −6 dB) were tested with three of the nine listeners. All

of the conditions are shown in Table 1. For each experimental condition and each listener, d′ scores were computed based on three blocks of 136 trials (i.e., 68 pairs of identical modulation patterns and 68 pairs of different modulation patterns, all pairs being presented in random order within a block).

Noise reduction algorithm

The NR algorithm used to process stimuli was based on a spectral subtraction technique representative of current hearing aid technology. The effects of this algorithm on speech perception have been investigated in previous studies (e.g., Sarampalis et al. 2009). The algorithm is described in detail by Fang and Nilsson (2004). It uses a single microphone input and estimates the noise level as the long-term stimulus average. The signal would then be any part of the input that is greater than this long-term average. However, in the current set of experiments, the NR algorithm was fed with separate, clean versions of the noise, together with a composite signal comprising both the signal and noise. Thus, this represented a perfect estimate of the continuous background noise. The average power level of the noise was estimated in nine frequency bands with center frequencies of 500, 750, 1,000, 1,500, 2,000, 3,000, 4,000, 6,000, and 8,000 Hz. As the bandwidth of the stimuli was relatively narrow, only the 1,000-Hz channel contained non-negligible amounts of energy. The bandwidth (−3 dB) of the 1,000-Hz channel was 360 Hz (cutoff frequencies: 865–1,225 Hz) and the attenuation rate outside the passband was 60 dB/octave. The instantaneous (averaged over 25 ms) power level of the input signal (i.e., the composite signal of both the signal and noise) was estimated for the same nine frequency bands. This instantaneous level was compared to the average level of the noise-alone stimulus

TABLE 1

Specification of the experimental conditions

	AM			FM	
Condition	SNR	NR	Condition	SNR	NR
1	+90[a]	Off	10	+90[a]	Off
2	+12	Off	11	+12	Off
3	+6	Off	12	+6	Off
4	0	Off	13	0	Off
5	−6	Off	14	−6	Off
6	+12	On	15	+12	On
7	+6	On	16	+6	On
8	0	On	17	0	On
9	−6	On	18	−6	On

Nine listeners completed fourteen conditions (no. 1–4, 6–8, 10–13, and 15–17), and three of the nine completed an additional four conditions (no. 5, 9, 14, and 18)

[a]The +90 dB SNR condition aimed to approach the "in quiet" condition

and a running SNR was computed within each frequency channel. A running attenuation, dependent on the SNR, was then applied to each channel. The relationship between attenuation and SNR was a broken stick function. For SNRs between 18 and 9 dB, attenuation increased linearly from 0 to 7 dB (slope of 0.78). For SNRs between 9 and −2 dB, attenuation increased linearly from 7 to 24 dB (slope of 1.55). The equation for the gain function is shown in Eq. 3. The NR algorithm was set to have a maximum attenuation of 24 dB and used the highest level of noise reduction (setting 3, as specified in Fang and Nilsson 2004).

$$\text{Attenuation} = \begin{cases} -1.55 \text{ SNR} + 20.91; & 9 \text{ dB} < \text{SNR} < 18 \text{ dB} \\ -0.78 \text{ SNR} + 14; & -2 \text{ dB} < \text{SNR} < 9 \text{ dB} \\ 24; & \text{SNR} < -2 \text{ dB} \end{cases}$$

(3)

As we were attempting to measure the peak performance of the NR algorithm, the model assumed a perfect knowledge of the noise. This signal known exactly (SKE) approach has a long standing acceptance in both engineering and perceptual studies (e.g., Dubbelboer and Houtgast 2007; Jørgensen and Dau 2011). In that SKE is taken to represent optimal performance, it is meaningful in the present context to demonstrate any notable limitation of the NR algorithm.

Testing procedure

Each condition was run three times as separate blocks with block order randomized within and across listeners. Each block consisted of 136 trials (i.e., one presentation of every modulation pattern pair) and typically lasted between 20 and 25 min. Listeners were given as much time as they needed to respond: typically they would run about four blocks in a 2-h session and complete the experiment in approximately 10 to 14 sessions.

RESULTS

Psychophysical data

Figure 1 shows the mean AM and FM pattern discrimination performance in terms of d' for the nine NH listeners as a function of SNR. The black markers show the results for the AM conditions, and the gray markers are for the FM conditions. For conditions without NR processing, the results are indicated by unfilled circles and for conditions with NR processing, they are indicated by asterisks. The quiet conditions run without masking are shown by filled squares.

Figure 1 shows that AM and FM mean discrimination scores were comparable when measured in quiet, in both cases with d' approximately equal to 2.0. Mean discrimination scores across listeners measured with
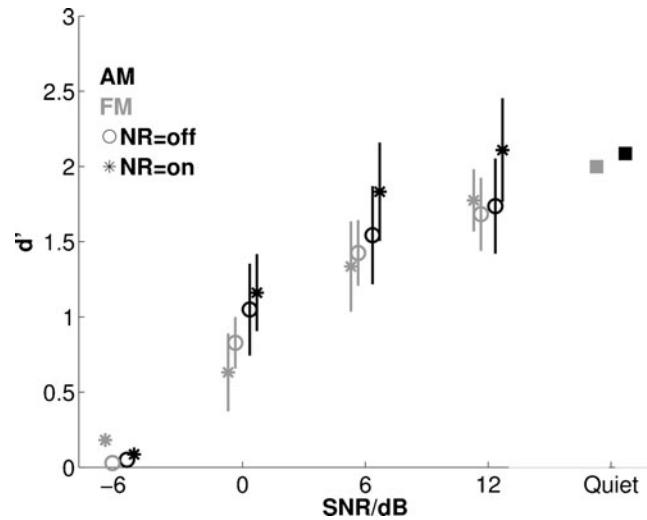

FIG. 1. Average discrimination scores for nine listeners (only three listeners were measured at an SNR of −6 dB). Noise reduction was found to yield small but systematic improvements in discrimination at SNRs of +6 and +12 dB for the AM conditions. For the FM conditions, NR had little effect, if any, on discrimination. The error bars represent the 95 % confidence intervals.

or without NR decreased similarly as a function of noise level for both AM and FM stimuli, and performance reached chance level (i.e., d' equal to 0.0) at an SNR of −6 dB. On average, NR improved AM discrimination slightly at positive SNRs (+6 and +12 dB) but had little effect, if any, at a 0-dB SNR. More precisely, the mean d' score increased from 1.54 to 1.83 at an SNR of +6 dB and from 1.74 to 2.11 at an SNR of +12 dB. Figure 1 also shows that at each SNR, NR did not noticeably affect FM discrimination.

A repeated measures analysis of variance (ANOVA) conducted on the d' scores confirmed the visual impressions. The ANOVA showed a significant effect for the following three main effects: SNR [$F(2,16)=$ 81.9, $p<0.0001$, $\varepsilon=0.91$, $\eta_p^2=0.91$]; noise reduction [$F(1,8)=195$, $p<0.0001$, $\varepsilon=1$, $\eta_p^2=0.96$]; modulation type (AM or FM) [$F(1,8)=8.97$, $p=0.017$, $\varepsilon=1$, $\eta_p^2=0.53$]. The ANOVA also showed a significant interaction between factors modulation type and noise reduction [$F(1,8)=7.83$, $p=0.023$, $\varepsilon=1$, $\eta_p^2=0.49$], and noise reduction and SNR [$F(2,16)=27.6$, $p<0.001$, $\varepsilon=0.52$, $\eta_p^2=0.78$]. A Student's t test on the means showed that there was a significant difference between the "NR on" and "NR off" conditions for AM at a 6-dB SNR ($p=0.03$) and 12-dB SNR ($p<0.001$). The ANOVA also showed a significant difference between the "NR on" and "NR off" conditions for FM at the 0-dB SNR ($p=0.04$), suggesting the NR tended to degrade FM discrimination at very low SNRs.

Important effects of NR not seen in averaged data were obtained for some listeners. Table 2 shows the individual discrimination scores for each experimen-

TABLE 2

Individual discrimination (d') scores for the nine listeners, together with the mean value (rightmost column)

	Subject									
Condition	1	2	3	4	5	6	7	8	9	Mean
AM quiet	2.25	1.78	1.47	2.01	2.23	2.52	2.22	2.18	2.11	2.09
AM +12 dB off	2.11	1.44	1.00	2.47	1.73	1.58	1.71	1.68	1.91	1.74
AM +12 dB on	2.30	1.61	1.53	3.08	1.98	2.14	1.98	2.19	2.18	2.11
AM +6 dB off	2.23	1.47	0.99	2.03	1.19	1.20	1.59	1.30	1.89	1.54
AM +6 dB on	2.02	1.45	0.99	2.21	1.53	1.90	2.00	2.12	2.28	1.83
AM 0 dB off	1.31	1.06	0.81	1.73	0.52	0.83	1.06	0.64	1.48	1.05
AM 0 dB on	1.10	0.84	0.60	1.74	1.16	1.15	1.32	1.07	1.47	1.16
AM −6 dB off	–	0.26	−0.33	0.23	–	–	–	–	–	0.05
AM −6 dB on	–	0.30	−0.19	0.15	–	–	–	–	–	0.09
FM quiet	2.15	1.69	1.52	2.48	2.38	2.31	1.89	1.57	2.00	2.00
FM +12 dB off	1.85	1.68	1.12	1.64	1.59	1.72	1.78	1.46	2.30	1.68
FM +12 dB on	1.59	1.48	1.37	1.92	2.05	1.78	1.72	1.86	2.21	1.78
FM +6 dB off	1.62	1.09	0.91	1.61	1.79	1.57	1.49	1.24	1.51	1.43
FM +6 dB on	0.84	0.96	0.94	1.75	1.76	1.31	1.12	1.52	1.83	1.34
FM 0 dB off	1.01	0.56	0.54	0.80	1.12	0.96	0.59	0.86	1.04	0.83
FM 0 dB on	0.35	0.12	0.24	0.88	1.08	0.74	0.54	0.96	0.78	0.63
FM −6 dB off	–	−0.12	−0.15	0.19	–	–	–	–	–	−0.03
FM −6 dB on	–	0.65	0.05	−0.15	–	–	–	–	–	0.18

tal condition. Indeed, large improvements (e.g., an effect size of more than 0.3) in AM discrimination scores were caused by NR for four of the nine listeners (no. 3, 4, 6, and 8) at +12 dB SNR and five of the nine listeners (no. 5, 6, 7, 8, and 9) at +6 dB SNR. For FM discrimination at 0 dB SNR, a relatively large decrease (e.g., an effect size of more than 0.3) in discrimination scores was caused by NR for three of the nine listeners (no. 1, 2, 3).

Modeling

The stimuli were processed with a model of early auditory processing (Patterson et al. 1987) to quantify the fidelity of AM and FM transmission (Sheft et al. 2008). This model was split into separate AM and FM processing routes. For both processing routes, the stimuli were presented to the model as in the experimental conditions: a composite stimulus consisting of the modulated signal plus the noise and a separate, stimulus consisting of the noise alone without the modulated signal. In addition to the experimental stimuli, a white noise component was added to the stimuli to simulate the variance in the performance of a listener. The amount of noise was set so that the predictions of the model matched the in-quiet conditions of the psychophysical data.

AM processing route

In the two stimulus intervals composing a given trial, AM information in the response to either the AM or FM stimuli was extracted at the output of a gamma-

tone filter tuned to either 850 or 1,000 Hz. The AM stimuli evoked dynamic variations in excitation level (i.e., temporal envelope fluctuations) at the output of the gammatone filter. For FM stimuli, AM information was consequent to the so-called "FM-to-AM conversion" (Saberi and Hafter 1995) occurring at the output of the gammatone filter. More precisely, the differential attenuation of cochlear filtering converted the frequency excursions of FM into dynamic variations in excitation level, in other words, into envelope fluctuations of the filter output. For the FM stimuli, an additional condition with a gammatone filter tuned to 850 Hz was also used to observe off-frequency FM-to-AM conversion (preliminary simulations were run for a wide range of center frequencies around 1 kHz, and these showed that maximum FM-to-AM conversion was obtained for a filter tuned to 850 Hz). For both the AM and FM stimuli, the AM information was determined by taking the absolute value of the Hilbert transform of the filtered signal. The resulting Hilbert envelope was then low-pass-filtered at 64 Hz using a first-order, zero-phase Butterworth filter (the envelope was passed forward and backward within the filter to compensate for the delay introduced by low-pass filtering). In each case (i.e., for AM or FM stimuli) and for each trial, the correlation of AM patterns extracted for the two stimulus intervals was calculated.

FM-processing route

The FM processing route was assessed only for the FM stimuli. In the two stimulus intervals composing a given trial, FM information was extracted at the

output of the gammatone filter tuned to 1 kHz with the correlation of the two extracted FM patterns calculated. The FM information was determined by taking the cosine of the angle of the Hilbert transform of the signal. For each trial, the correlation of FM patterns extracted for the two stimulus intervals was calculated. AM-to-FM conversion was not included. Heinz and Swaminathan (2009) showed that AM-to-FM conversion is extremely weak and as such probably would not contribute to speech perception (see Fig. 7B, J. Assoc. Res. Otolaryngology, 10, 407-423, 2009). Heinz and Swaminathan calculated neural cross-correlation coefficients of spike trains from an auditory nerve model for different speech stimuli. They showed there was very low correlation of temporal fine structure information between stimuli with a speech envelope and a random fine structure and the original speech stimuli. This shows that there is no useful temporal fine structure information introduced into the stimuli by an AM-to-FM conversion process taking place at the output of auditory filters.

For each experimental condition, average AM and FM correlations were calculated by taking the arithmetic means of the correlations across the 68 "same" and 68 "different" trials. These average correlations were determined for eight SNRs (–6, –3, 0, +3, +6, +12, +24, and +90 dB), with and without noise reduction. The correlation values, r, were transformed into Z-scores using the Fisher transform (Fisher 1915) as shown in Eq. 4.

$$Z = \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) \qquad (4)$$

From these Z-scores of the same (Z_{same}) and different (Z_{diff}) trials, for a particular condition, a d′ score was determined as shown by Eq. 5. Eight repetitions were used for each condition.

$$d' = \frac{\bar{Z}_{\text{same}} - \bar{Z}_{\text{diff}}}{\sqrt{\sigma_{Z_{\text{same}}}{}^2 + \sigma_{Z_{\text{diff}}}{}^2}} \qquad (5)$$

Figure 2 shows the output of the model. In Figure 2A, the fidelity of AM transmission is shown in response to an AM input. The solid line shows the d′ predictions for stimuli that have not been processed with the NR algorithm, while the dashed line indicates model performance for stimuli that have undergone NR. For the "in-quiet" condition, without NR, d′ is just above 2. As the SNR is decreased to 3 dB, d′ falls slightly to 1.7, and as the SNR is reduced to –6 dB, d′ falls more sharply to 0.5. The introduction of NR does not change model performance for the in-quiet condition or with an SNR of 24 dB. As the SNR is reduced to 6 dB, d′ increases slightly to 2.3. A further decrease in SNR to –6 dB causes d′ to fall sharply to
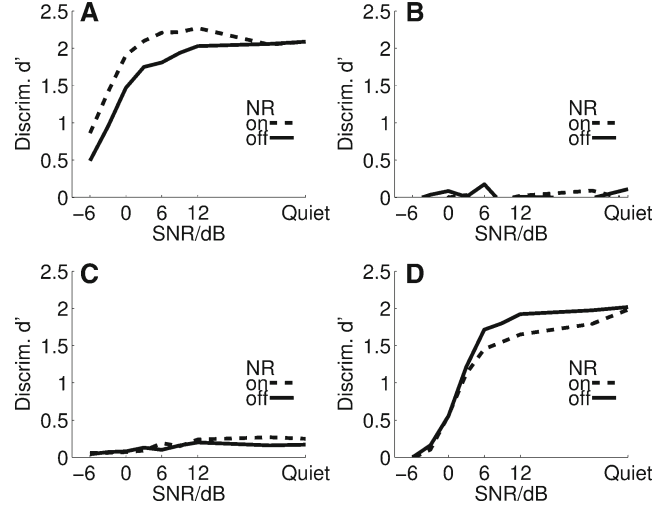

FIG. 2. Model predictions showing the effects of NR on the fidelity of transmission for **A** AM information (input: AM stimuli), **B** on-frequency AM information at 1 kHz, converted from FM information (input: FM stimuli), **C** off-frequency AM information at 850 Hz, converted from FM information (input: FM stimuli), and **D** FM information (input: FM stimuli). The *solid lines* show the predictions without noise reduction, and the *dashed lines* show the predictions with noise reduction.

just below 1. Overall, the model shows that the fidelity of AM transmission is improved by the NR system for SNRs of 12 dB and less.

Figure 2B shows the fidelity of the on-frequency AM transmission in response to an FM input (i.e., FM converted into AM as a result of the differential attenuation of cochlear filtering using a gammatone filter tuned to 1 kHz). This plot shows that either with or without NR, there is no resulting AM information that can be used to discriminate the patterns of modulation. Figure 2C shows the fidelity of off-frequency AM transmission in response to an FM input (i.e., FM converted into AM as a result of the differential attenuation of cochlear filtering using a gammatone filter tuned to 850 Hz; in a preliminary modeling study, we found that this filter produced the maximum FM-to-AM cues for a centre frequency of 1 kHz). Again, there is no resulting off-frequency AM information that can be used to discriminate the patterns of modulation. These results suggest that NR should have no effect on the discrimination of the FM stimuli if the cues used to discriminate modulation pattern pairs are based solely on AM cues resulting from the conversion of FM into fluctuations in excitation level at the output of cochlear filtering.

Figure 2D shows the fidelity of FM transmission for the FM stimuli. For the in-quiet condition, without NR, d′ has a value of 2. As the SNR is decreased to 6 dB, d′ gradually reduces to 1.7. For SNRs below 6 dB, d′ reduces much more steeply and has a value of 0 at an SNR of –6 dB. With NR on, the d′ value for the in-quiet condition is the same as obtained without NR.

A detrimental effect, however, of NR is observed for SNRs of 6 to 24 dB. At lower SNRs, d′ declines steeply irrespective of NR. These results suggest that when NR is on, there may be some small reduction in the discrimination of FM for SNRs between roughly 6 and 24 dB.

Comparison with model

The model predictions described in the "Modeling" section compared reasonably well to the psychophysical data reported in "Psychophysical data" section. For AM stimuli (Fig. 2A), the model correctly predicts that NR would improve discrimination for SNRs of +12 and +6 dB. Also, d′ values compare well to the in-quiet condition (although the model was actually tuned so this would necessarily be the case). As the SNR is reduced to 0 and −6 dB, the model over-estimates performance, i.e., d′ values for AM discrimination at SNRs of 0 and −6 dB are too high. For AM cues that are evoked by FM at the output of cochlear filters (Fig. 2B, C), the model predicts that useful cues are extremely weak regardless of whether the cochlear filter is on-frequency (Fig. 2B) or off-frequency (Fig. 2C). For "true" FM cues, i.e., those elicited by the FM pattern at the output of the gammatone filters (Fig. 2D), the model predicts a d′ score of 2 (with and without NR) for the "in-quiet" condition (the level of the internal noise was actually set so to match the in-quiet condition). As the SNR is reduced to 12 and 6 dB, the d′ scores calculated with and without NR begin to deviate with the model predicting poor discrimination performance with NR. This does not agree with the average psychophysical data and shows a limitation of the model for FM transmission. Further reductions in the SNR produce no difference between the predictions of d′ either with or without NR.

DISCUSSION

The present study showed that an NR algorithm based on spectral subtraction improved the discrimination of narrowband complex AM patterns. The same algorithm had little, if any, effect on the discrimination of complex narrowband FM patterns. Our findings complement previous studies (e.g., Dubbelboer and Houtgast 2007) by using non-linguistic stimuli. They also highlight the perceptual effects of the distortions produced by NR in the modulation domain.

The improvement in AM discrimination showed that NR could be potentially useful in enhancing AM features that were present in the original signal. A detailed inspection of stimuli waveforms suggested that the pairs of complex patterns that benefitted most from the NR algorithm were those that differed in their segments of highest amplitude (i.e., their main or primary envelope peaks). In other words, those high-amplitude features remained intact (or were enhanced) when the noise was removed. Also, those trials that differed in their segments of lowest amplitudes (i.e., their secondary envelope peaks) benefitted least from the NR algorithm. This is illustrated in Figure 3.

The left-hand panels of Figure 3 show the pairs of AM patterns, for a particular trial, that produce the largest increase in average percent correct scores when NR is switched off (upper left panel) compared to when it is switched on (lower left panel). This represents a trial for which NR was particularly beneficial and the increase in scores was from 45 % to 89 % correct. The increase in discrimination seems to come from the higher amplitude peaks becoming more prominent (and more discriminable) when NR
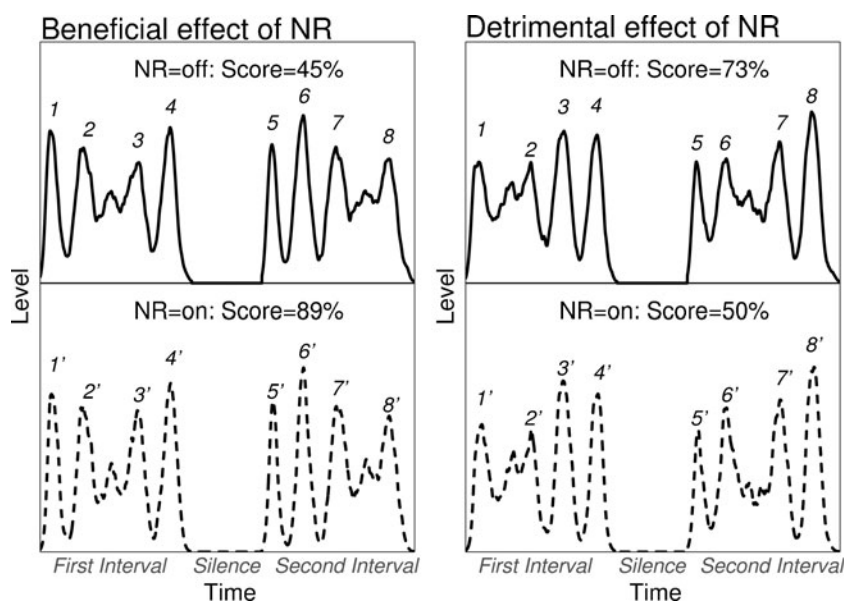

FIG. 3. Pairs of envelope patterns that produced the most beneficial (left panel) and the most detrimental (right panel) effects of NR with an SNR of +12 dB. Each panel shows the AM patterns for a particular trial (i.e., the first and the second intervals separated by a silent interval). The top-left panel shows the pair of AM patterns that produced an average score of 45 %, with NR off and this increases to an average scores of 89 % with NR on (shown in the bottom-left panel). The top-right panel shows the pair of AM patterns that produced an average score of 73 %, with NR off and this decreases to an average scores of 50 % with NR on (shown in the bottom-right panel). *Numbers* indicate the peaks within the pattern of each trial, see text for details.

is switched on (compare peak 4 with 4′, peak 6 with 6′, and peak 7 with 7′). The correlation of the two intervals without NR was 0.71, and this decreases to a correlation value of 0.60 when NR is switched on (therefore discrimination is easier with NR).

The right-hand panels of Figure 3 show the pairs of AM patterns, for a particular trial, that produce the largest decrease in average percent correct scores when NR is switched on (lower right panel) compared to when it is switched off (upper right panel). This represents a trial for which NR was particularly detrimental, and the decrease in scores was from 73 % to 50 % correct. In this case, it is more difficult to see exactly where the decrease in performance arises. However, listeners would need to use the lower amplitude peaks (labeled 2 and 6 for the NR = off condition and 2′ and 6′ for the NR = on condition) to discriminate the two AM patterns as it is these peaks which are different in the two intervals. The correlation of the two intervals without NR was 0.53, and this increases slightly to a correlation value of 0.57 when NR is switched on.

Also, a small improvement for AM discrimination was, in part, due to the restoration of the similarity of those trials that were the same (before the addition of the noise); consequently, a listener could be more confident that an interval contained two identical patterns of modulation and would produce fewer errors. This is supported by a small increase in the correct rejection rate for AM discrimination with NR at SNRs of 12 and 6 dB. Also, the improvement shown by the model, with NR at 12 and 6 dB for AM stimuli, is due to an increase in the correlation values of those trials that are the same. The NR algorithm worked best when the noise was relatively low in level compared to the signal and did not impair too many features of the signal. These results suggest that NR would be of limited benefit to the transmission of narrowband AM speech cues and thus to speech recognition when SNR falls below 6 dB.

The NR algorithm did not affect FM discrimination for narrowband signals unless the SNR was low. This is not surprising considering the mode of operation of the algorithm. The NR algorithm attenuates the signal by an amount that is inversely related to the SNR. For an SNR of approximately +18 dB, the attenuation applied by the algorithm is 0 dB. This increases to a maximum attenuation of 24 dB (i.e., a gain of −24 dB) for SNR values of −2 dB and below. The peak-to-trough ratio of the amplitude envelope is (by definition) much lower for the FM than AM signal. Therefore, the local SNR (defined as the running average of a 25-ms window) will vary much more for the noisy AM signal than for the noisy FM signal. In fact, for the FM signal at the higher SNRs, the local SNR will be too high to be attenuated by the NR

algorithm. This is why the NR algorithm did not produce any change in discrimination of the FM patterns. It is only when the SNR was reduced, such that the variations in the amplitude of the noise contributed significantly to the overall variation in the amplitude of the combined signal and noise that the NR algorithm could operate and apply attenuation. However, when such low SNRs were reached, the discrimination of the FM patterns became difficult, and the NR algorithm had little effect. Also, at these low SNRs, the NR algorithm was most likely removing much of the signal in addition to the noise, particularly because the signal was narrowband.

It should be noted that our approach of using an SKE method (i.e., having an exact knowledge of the noise) may afford better results from the NR algorithm compared to real-life situations. For example, an NR algorithm working on a mixed signal of noise and speech might introduce different distortions of the AM and FM information. This is important to consider when extrapolating the results of the present study to real applications of NR. Also, all the stimuli in the current study were narrowband, and as such, we do not observe any effects regarding the operation of the NR algorithm on more realistic, broadband stimuli such as speech. For example, with broadband stimuli, one might expect the NR algorithm to attenuate portions in each channel where the signal level is low and that this might then enhance spectral contrasts.

Previous studies showed that NR algorithms based on spectral subtraction fail to improve speech intelligibility (e.g., Hu and Loizou 2007; Sarampalis et al. 2009). However, several low-level (i.e., sensory) and high-level (e.g., linguistic) effects of NR may underlie the lack of speech perception benefits. Overall, our results (based on non-linguistic stimuli) suggest that such NR algorithms do not alter the transmission of narrowband FM cues and may potentially improve the transmission of narrowband AM cues. Still, the observed benefits were restricted to high SNRs and were generally small. These psychophysical results suggest that the lack of benefits of this NR algorithm on speech intelligibility is partly caused by its limited effect on the transmission of narrowband speech modulation cues. Nevertheless, detailed analysis of individual data revealed that some listeners were more sensitive to the effects of NR (both positive and negative) than others. NH listeners may vary largely in terms of their suprathreshold modulation processing capacities, which is consistent with recent work by Ruggles et al. (2011). This also suggests that a customized approach to the prescription of NR algorithms may be beneficial to the user (e.g., Arehart et al. 2007).

CONCLUSIONS

The current study investigated the effect of a single microphone noise reduction algorithm based on spectral subtraction on the transmission of narrow-band AM and FM information. In addition to measuring the psychophysical performance when stimuli were processed by the NR algorithm, the transmission of the AM and FM information using a model of early auditory processing was also assessed. The psychophysical and modeling study showed that:

1. The NR algorithm yielded a small but significant increase in NH listeners' ability to discriminate complex AM patterns for SNRs of +6 and +12 dB.
2. Overall, the NR algorithm did not affect discriminability of complex FM patterns. Any degradation was limited to an SNR of 0 dB.
3. The effect of NR on the transmission of the complex AM patterns could largely be predicted by a simple model of early auditory processing comparing modulation patterns at the output of cochlear filters. This suggests that the small beneficial effects of NR on the transmission of temporal envelope cues were mainly constrained by low-level auditory processes.
4. With regard to the transmission of the complex FM patterns, the model predicted a detrimental effect of NR for strictly positive SNRs, in contrast to the average psychophysical data.

With modulation cues central to speech perception, the limited effects of NR on modulation perception are consistent with the limited effects of NR on speech recognition reported in previous studies. However, individual differences suggest that NR may be either beneficial or detrimental for at least some listeners.

ACKNOWLEDGMENTS

REFERENCES

ARDOINT M, LORENZI C, PRESSNITZER D, GOREA A (2008) Perceptual constancy in the temporal envelope domain. J Acoust Soc Am 123:1591–1601

AREHART KH, KATES JM, ANDERSON MC, HARVEY LO (2007) Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners. J Acoust Soc Am 122:1150–1164

DARWIN CJ (2008) Listening to speech in the presence of other sounds. Phil Trans R Soc B 363:1011–1021

DUBBELBOER F, HOUTGAST T (2007) A detailed study on the effects of noise on speech intelligibility. J Acoust Soc Am 122:2865–2871

EPHRAIM Y, MALAH D (1984) Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics Speech and Signal Processing 32:1109–1121

EPHRAIM Y, MALAH D (1985) Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE Transactions on Acoustics Speech and Signal Processing 33:443–445

FANG X, NILSSON MJ (2004) Noise reduction apparatus and method. US Patent US-6757395-B1

FISHER RA (1915) Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. Biometrika (Biometrika Trust) 10(4):507–521

FRENCH NR, STEINBERG JC (1947) Factors governing the intelligibility of speech sounds. J Acoust Soc Am 19:90–119

HEINZ MG, SWAMINATHAN J (2009) Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech. J Assoc Res Otolaryngol 10(3):407–423

HOUTGAST T, STEENEKEN HJM (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

HU Y, LOIZOU PC (2007) A comparative intelligibility study of single-microphone noise reduction algorithms. J Acoust Soc Am 122:1777–1786

IVES DT, LORENZI C (2011) Perception of time-compressed modulation patterns, 34th Annual Midwinter meeting of the Association for research in Otolaryngology

JØRGENSEN S, DAU T (2011) Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing. J Acoust Soc Am 130:1475–1487

MACMILLAN NA, CREELMAN CD (2005) Detection theory: A user's guide, 2nd edn. Lawrence Erlbaum, Mahwah Associates

MOORE BCJ (2008) Basic auditory processes involved in the analysis of speech sounds. Phil Trans R Soc B 363:947–963

PATTERSON RD, NIMMO-SMITH I, HOLDSWORTH J, RICE P (1987) An efficient auditory filterbank based on the gammatone function. Meeting of the IOC Speech Group on Auditory Modeling at RSRE. Malvern, England

RUGGLES D, BHARADWAJ H, SHINN-CUNNINGHAM BG (2011) Normal hearing is not enough to guarantee robust encoding of supra-threshold features important in everyday communication. PNAS 108(37):15516–15521

SABERI K, HAFTER ER (1995) A common neural code for frequency- and amplitude-modulated sounds. Nature 374:537–539

SARAMPALIS A, KALLURI S, EDWARDS B, HAFTER E (2009) Objective measures of listening effort: effects of background noise and noise reduction. J Sp Lan & Hear Res 52:1230–1240

SHANNON RV, ZENG FG, KAMATH V, WYGONSKI J, EKELID M (1995) Speech recognition with primarily temporal cues. Science 270:303–304

SHEFT S, LORENZI C (2008) Discrimination of stochastic patterns of frequency modulation relevant to speech perception. J Acoust Soc Am 123:3711

SHEFT S, ARDOINT M, LORENZI C (2008) Speech identification based on temporal fine structure cues. J Acoust Soc Am 124:562–575

SHEFT S, SHAFIRO V, LORENZI C, MCMULLEN R, FARRELL C (2012) Effect of age and hearing loss on the relationship between stochastic FM discrimination and speech perception. Ear Hear 33(6):709–720

STEENEKEN HJ, HOUTGAST T (1980) A physical method for measuring speech-transmission quality. J Acoust Soc Am. Jan 67(1):318–26

ZENG FG, NIE K, STICKNEY GS, KONG YY, VONGPHOE M, BHARGAVE A, WEI C, CAO K (2005) Speech recognition with amplitude and frequency modulations. Proc Natl Acad Sci U S A 102:2293–2298