Contribution of envelope periodicity to release from speech-on-speech masking

Claus Christiansen,[a] Ewen N. MacDonald, and Torsten Dau
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Building 352, DK-2800 Kgs. Lyngby, Denmark

(Received 19 September 2012; revised 13 May 2013; accepted 27 June 2013)

Masking release (MR) is the improvement in speech intelligibility for a fluctuating interferer compared to stationary noise. Reduction in MR due to vocoder processing is usually linked to distortions in the temporal fine structure of the stimuli and a corresponding reduction in the fundamental frequency (*F*0) cues. However, it is unclear if envelope periodicity related to *F*0, produced by the interaction between unresolved harmonics, contributes to MR. In the present study, MR was determined from speech reception thresholds measured in the presence of stationary speech-shaped noise and a competing talker. Two types of processing were applied to the stimuli: (1) An amplitude- and frequency-modulated vocoder attenuated the envelope periodicity and (2) high-pass (HP) filtering (cutoff = 500 Hz) reduced the influence of *F*0-related information from low-order resolved harmonics. When applied individually, MR was unaffected by HP filtering, but slightly reduced when envelope periodicity was attenuated. When both were applied, MR was strongly reduced. Thus, the results indicate that *F*0-related information is crucial for MR, but that it is less important whether the *F*0-related information is conveyed by low-order resolved harmonics or by envelope periodicity as a result of unresolved harmonics. Further, envelope periodicity contributes substantially to MR.
 [http://dx.doi.org/10.1121/1.4816409]

PACS number(s): 43.71.Es, 43.71.Gv, 43.71.An, 43.71.Bp [JFC] Pages: 2197–2204

I. INTRODUCTION

The most important mode of communication in our daily life is speech. However, in many situations, speech communication takes place in adverse conditions with high levels of background noise, several interfering talkers and/or reverberation. Normal-hearing (NH) listeners are typically able to understand speech even at very low signal-to-noise ratios (SNRs). In conditions where the interfering sound is amplitude modulated noise or a competing talker, NH listeners are commonly able to utilize speech information in the low-amplitude parts of the interferer such that they are able to understand the speech at a much lower SNR than in the case of a stationary-noise interferer. This ability has usually been referred to as "listening-in-the-dips" and the corresponding improvement in speech intelligibility has been termed masking release (MR). Compared to NH listeners, hearing-impaired (HI) listeners often need much higher SNRs to understand speech in noise and often show very little or no MR (e.g., Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Lorenzi *et al.*, 2006; Bernstein and Grant, 2009). Even after compensating for reduced sensitivity with hearing aids, many HI listeners still show great difficulties in adverse listening conditions (e.g., Duquesnoy and Plomp, 1983; Gustafsson and Arlinger, 1994; Metselaar *et al.*, 2008). The reduced MR experienced by HI listeners has traditionally been ascribed to reduced frequency selectivity or an increased amount of forward masking that might limit their ability to benefit from spectral and temporal dips in the masker (e.g., Festen and Plomp, 1990; Baer and Moore, 1994; Dubno *et al.*, 2003). Furthermore, it has been proposed that deficits in the processing of the temporal fine structure of the stimuli affect the coding of the stimuli's fundamental frequency (*F*0) in HI listeners (e.g., Qin and Oxenham, 2003; Hopkins *et al.*, 2008; Oxenham and Simonson, 2009) or in auditory aging (e.g., Pichora-Fuller *et al.*, 2007; MacDonald *et al.*, 2010).

Coding of *F*0 plays an important role for the perceptual segregation of concurrent and sequential sources (Brokx and Nooteboom, 1982; Darwin, 1997) and may be an important factor in MR observed when the masker is a competing talker (e.g., Qin and Oxenham, 2003; Bernstein and Grant, 2009; Bernstein and Brungart, 2011; Christiansen and Dau, 2012). In general, there are two different theoretical concepts describing how the *F*0 of a stimulus can be extracted by the auditory system: Place or temporal coding. In terms of place coding (pattern matching), the *F*0 of a stimulus can be extracted by matching harmonic templates to the basilar membrane (BM) excitation pattern (e.g., Wightman, 1973; Cohen *et al.*, 1995). In terms of temporal coding, the firing of auditory-nerve cells synchronous with BM vibration can be used to extract the *F*0 of the input stimulus via inter-spike intervals (ISIs; e.g., Licklider, 1951; de Cheveigné, 1998). At low frequencies, the spectral harmonics of voiced speech are spatially resolved by the auditory system and the *F*0 of the input stimuli can be extracted from the BM excitation pattern or from the period of individual frequency components in the corresponding channels. At high frequencies, the harmonics are considered to be spatially unresolved due to the increasing bandwidth of the auditory filters with increasing center frequency. However, interaction between harmonics

[a]Author to whom correspondence should be addressed. Electronic mail: cfj@oticon.dk

within the same auditory filter gives rise to envelope periodicity related to the *F*0 of the stimuli. Since the ability of auditory-nerve cells to phase lock to the vibration of the BM is progressively reduced for increasing frequency, it is generally assumed that at high frequencies, the ISIs of the auditory nerve cells reflect the periodicity of the envelope fluctuations (e.g., Palmer and Russell, 1986).

Several studies have shown that low-order harmonics provide better *F*0 discrimination performance than high-order unresolved harmonics (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994) and dominate the perceived *F*0 in the case of conflicting cues (e.g., Plomp, 1967; Micheyl and Oxenham, 2007). However, the contribution of envelope periodicity to MR has not been investigated explicitly.

In order to examine the importance of low-frequency *F*0 information for MR, Qin and Oxenham (2006) measured the benefit of adding unprocessed low-frequency information to envelope-vocoder processed speech. They used an eight-channel noise-excited vocoder where unprocessed low-frequency information below 300 or 600 Hz was added by replacing the lower frequency channels with a low-pass (LP) filtered version of the original stimulus. The speech intelligibility was measured with steady-state speech-shaped noise and a competing talker. The results of Qin and Oxenham (2006) showed that while adding unprocessed low-frequency information significantly increased speech intelligibility, listeners never reached the same level of performance achieved with the unprocessed stimuli. A particularly interesting result was that even with unprocessed low-frequency information up to 600 Hz the MR was very close zero. In contrast, the MR obtained with unprocessed stimuli was about 7 dB. This large difference in MR might have been caused by distortions of the envelope periodicity of the higher-frequency channels, introduced by the noise vocoding. These results suggest that envelope periodicity might be important for MR.

In a later study, Oxenham and Simonson (2009) investigated if pitch information provided by the low-order resolved harmonics is important for MR. They measured MR for NH listeners using LP and high-pass (HP) filtered stimuli in order to either retain or eliminate low-order harmonics, while achieving the same speech intelligibility in steady-state noise. In both conditions, MR was greatly reduced. Oxenham and Simonson (2009) suggested that MR might be determined mainly by the perceptual redundancy of the target speech instead of the *F*0 of resolved harmonics.

Stone *et al*. (2008) investigated the role of envelope periodicity for speech perception using vocoder processing and found that NH listeners showed a speech intelligibility benefit with an interfering talker. Further, a difference in speech perception was observed across vocoding strategies suggesting that envelope periodicity cues may be stronger when presented using a tone- as opposed to a noise vocoder. In contrast, Xu and Zheng (2007) found that NH listeners showed no speech intelligibility benefit from envelope periodicity in stationary noise. However, Xu and Zheng measured vowel and consonant recognition whereas envelope periodicity may be more important for speech intelligibility, which is a sequential rather than simultaneous task. Taken together, these results support the indications of Qin and Oxenham (2006) that envelope periodicity may contribute substantially to MR, particularly in a speech-on-speech context.

There are two ways in which identifying the *F*0 of a voiced speech target could be important for MR in speech-on-speech situations: Through identifying the time intervals that contain target speech versus those that contain competing speech, and through identifying the (audio) frequency regions that contain target speech versus those that contain competing speech. At higher frequencies, this information may be provided by the envelope periodicity produced by unresolved harmonics.

To test if envelope periodicity contributes to MR, the present study used a novel signal processing technique to attenuate this periodicity. The technique was based on a traditional tone-vocoder but maintains the instantaneous frequency (IF) in each channel. Briefly, the input signal was divided into 16 frequency channels for which both the amplitude and the IF course were estimated. In each channel, a LP filtered version of the IF course was used to drive a sine generator which was then modulated by a LP filtered version of the estimated envelope. Finally, all channels were recombined to generate the output signal. Using frequency-modulated tone carriers, a natural sounding speech output could be obtained using a relatively small number of channels. This allowed for a relatively large separation of carrier frequencies in adjacent channels so that envelope periodicity was not reintroduced in each channel due to interaction between carriers when the channels were recombined. The attenuation of envelope periodicity was combined with a reduction in the *F*0-related information from low-order resolved harmonics obtained via HP filtering with a relatively low cut-off frequency of 500 Hz. It was thus possible to investigate separately the effect of reduced resolved harmonics and reduced envelope periodicity, as well as the effect of reducing both.

In the present study, speech intelligibility was measured in four different processing conditions. The cutoff-frequency of the envelope LP filter was chosen to be either 30 or 300 Hz in order to attenuate or retain envelope periodicity. After vocoding, the stimuli were either HP filtered at 500 Hz or left unprocessed in order to either reduce or retain *F*0-related information conveyed by the low-order resolved harmonics.

II. SIGNAL PROCESSING

The stimuli were processed by a vocoder with amplitude and frequency modulated tone carriers, as illustrated in Fig. 1. First, the signal was decomposed into 16 frequency channels using a gammatone filterbank (Patterson *et al*., 1987). The filterbank consisted of fourth-order gammatone bandpass filters with center frequencies ranging from 50 to 7500 Hz, equally spaced on an equivalent-rectangular-bandwidth number scale (ERB_N; Glasberg and Moore, 1990), each with a bandwidth of 1 ERB_N. In each channel, the envelope and IF were estimated in two parallel paths. The envelope was calculated by the absolute value of the analytical signal (via the Hilbert transform) and LP filtered with a

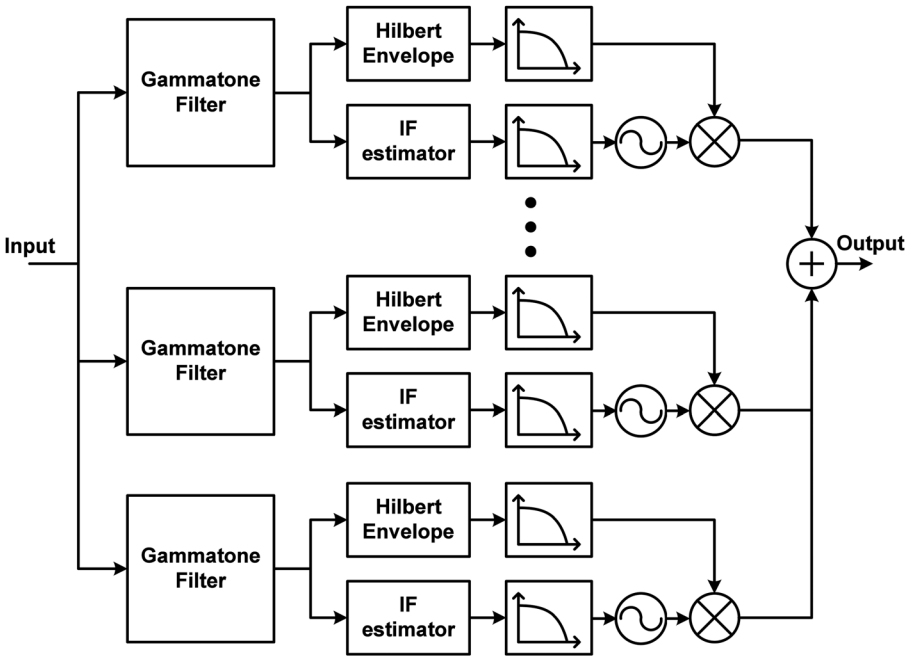

FIG. 1. Schematic of the amplitude and frequency modulated vocoder. The input signal is divided into 16 frequency channels. In each channel the envelope and IF is calculated and LP filtered. The IF drives a sine generator where the amplitude is modulated by the estimated envelope.

fourth-order Butterworth filter (24 dB/octave slope). The IF was estimated using the algorithm described in Nguyen *et al.* (2009), which is a Kalman smoother based dynamic autoregressive model developed for tracking the IF of noisy and non-stationary sinusoids. The estimated IF was smoothed with a 50-ms median filter and a 50-ms moving-average filter. The estimated IF was used to drive a sine generator and the output signal was amplitude modulated by the envelope signal. Before recombining all the channels, the root-mean-square (rms) value in each channel was normalized to the input rms in the corresponding channel. The final signal was scaled to have the same overall level as the input to the vocoder.

For low frequency channels, the IF driving the tone vocoder was usually a harmonic of $F0$ of the speech input. However, for higher frequency channels, the IF was not usually harmonically related to $F0$. For example, the IF in channels with the center frequencies of 119 and 210 Hz, respectively, was within 10 Hz of a harmonic of $F0$ 67% and 57% of the time, respectively. Conversely, the IF in all of the 9 frequency channels above 1000 Hz was within 10 Hz of a harmonic of $F0$ less than 20% of the time. Note, however, that any change in harmonicity introduced by this vocoder is the same regardless of whether the envelopes are LP filtered at 30 or 300 Hz.

In order to remove or retain $F0$-related envelope modulations, the cut-off frequency of the envelope LP filter was either 30 or 300 Hz. The speech and interferer were processed independently and mixed after the processing. The mixture was either presented to the listeners without any further processing or HP filtered with a cut-off frequency of 500 Hz. Thus, four different conditions were considered in the experiment: Two broadband conditions with a 30-Hz (BB30) or a 300-Hz (BB300) envelope filter in the vocoder, and two HP filtered conditions with a 30-Hz (HP30) or a 300-Hz (HP300) envelope filter in the vocoder.

The HP filtering procedure was conducted in the same manner as described by Oxenham and Simonson (2009). The signals were mixed at the appropriate SNR and then HP filtered at 500 Hz with a fourth-order Butterworth filter. An off-frequency masker was generated by LP filtering speech-shaped noise at 500 Hz (fourth-order Butterworth filter), and the rms level was adjusted to 12 dB below the level of the target sentence.

By using a vocoder with frequency modulated carriers it is possible, to some extent, to preserve the original temporal and spectral structure in the speech signal and obtain a natural-sounding representation of the speech using only 16 channels. In the 16-channel filterbank, the separation between the center frequencies is relatively large ($\approx$2 ERB_N). This is advantageous because it avoids interactions between the carriers when the channels are recombined. Thus, using this processing technique, it is possible to generate stimuli that have a similar overall spectro-temporal energy pattern but with reduced envelope periodicity. This is illustrated in Fig. 2, which shows the auditory spectrogram of a sentence processed by the vocoder with an envelope cut-off frequency of 30 Hz (left panel) and 300 Hz (right panel), respectively. The auditory spectrogram was produced using 128 fourth-order gammatone filters, ranging from 0 to 8000 Hz and equally spaced on an ERB_N number scale. The output of each channel is the Hilbert envelope LP filtered at 500 Hz. The two panels show that the overall spectro-temporal structure of the processed signals is very similar. The main difference is found in the higher frequency channels where envelope periodicity can be seen in the bottom panel (300-Hz envelope LP filter) but not in the top panel (30-Hz envelope LP filter).

Figure 3 shows a more detailed comparison of the envelopes obtained with the 30- and 300-Hz envelope cut-off frequency in the frequency channels at 119-, 1085-, and 2958-Hz, respectively. The envelopes shown in Fig. 3 are based on an analysis of the *processed* signal (i.e., the vocoder output), using the same gammatone filterbank and envelope extraction as in the vocoder. This was done to verify that

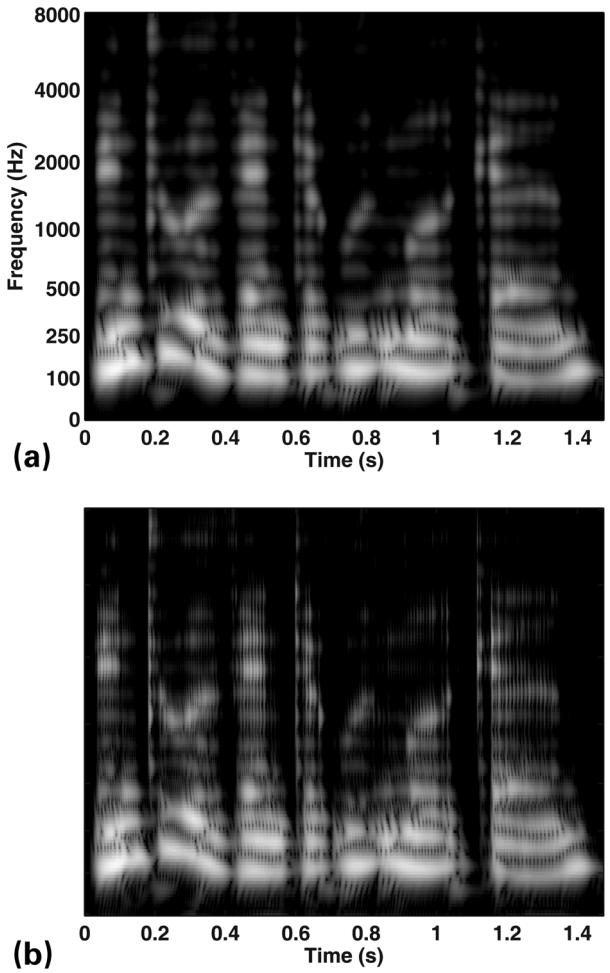

FIG. 2. Auditory spectrogram of the sentence ("Han hoppede op på cyklen"), processed by the vocoder with a 30-Hz (left panel) and a 300-Hz (right panel) envelope LP filter. The spectrograms were obtained using a 128-channel gammatone filterbank, with 1-ERB_N wide filters, equally spaced on an ERB_N number scale.

envelope periodicity was not reintroduced in the processed signal due to interaction between the carriers (e.g., Ghitza, 2001).

In the top panel of Fig. 3, it is clear that the two envelopes are almost identical due to the slow fluctuations in the amplitude of the signal. At 1085 Hz (the middle panel of Fig. 3), the 300-Hz condition shows a small amount of envelope periodicity related to the *F*0 of the talker, which is not reflected in the 30-Hz condition. Similarly, at 2958 Hz (the bottom panel of Fig. 3), no envelope periodicity is seen in the 30-Hz condition. However, in the 300-Hz condition, envelope periodicity is clearly visible. Thus, the reanalysis demonstrates that envelope periodicity in the mid- to high-frequency channels is attenuated or completely removed when the 30-Hz LP filter is applied to the envelope in the vocoder processing.

III. METHOD

A. Listeners

Five NH listeners with audiometric thresholds of 20 dB hearing level (HL) or less at all measured frequencies between 125 and 8000 Hz participated in the experiment. The age of the listeners ranged between 22 and 32 yrs with a mean age of 27.

B. Speech material

Speech reception thresholds (SRTs) were measured using the Danish speech intelligibility test called conversational language understanding evaluation (CLUE; Nielsen and Dau, 2009), which is very similar to the hearing-in-noise test originally developed for English (Nilsson *et al.*, 1994). The CLUE material consists of natural and meaningful sentences representing conversational speech and has a fixed structure consisting of five words per sentence. The sentences were spoken by a male talker with an average fundamental frequency (*F*0) of 119 Hz.

The maskers were an unintelligible single talker and a stationary speech-shaped noise (SSN). The single talker was the international speech test signal (ISTS; Holube *et al.*, 2010), which consists of natural speech from six female talkers speaking different languages that have been segmented and remixed using a randomization procedure in order to make it largely unintelligible. The average *F*0 of the ISTS signal was 207 Hz. The stationary noise was equalized to have the same long-term spectrum as the ISTS signal.

Only the target speech and the competing talker (ISTS) were processed by the vocoder to remove the envelope periodicity. The stationary noise was left unprocessed.

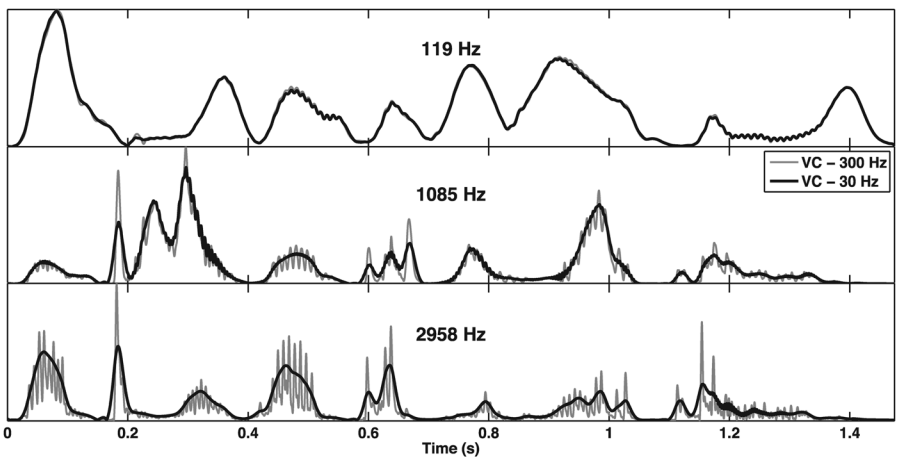

FIG. 3. Analysis of the envelopes of the vocoded signals. The processed signals were analyzed using the same front end as the vocoder (16-channel gammatone filterbank and LP filtered Hilbert envelopes). The three panels show the output of the channels with the center frequencies: 119-Hz (top), 1085-Hz (middle), and 2958-Hz (bottom). In each panel, the envelope of the vocoded signal with a 300-Hz (dark-gray) and a 30-Hz (black) LP filter is shown, respectively.

C. Procedure

The experiment was conducted in a double-walled sound insulated booth, where the experimenter controlled the procedure by means of a MATLAB application developed specifically for the CLUE test. The digital signals were sampled at 22 050 Hz and converted to analog signals by a high-end 24 bit soundcard (RME DIGI96/8, Haimhausen, Germany). The stimuli were presented diotically over Sennheiser (Wennebostel, Germany) HD580 headphones. The target sentences were presented at a fixed level of 65 dB sound pressure level (SPL), whereas the level of the interferer was determined via an adaptive procedure used to measure the SRTs. The onset and offset of the interferer were 1 s before and 600 ms after the sentence, respectively, and 400 ms ramped squared-cosine window was applied to the onset and the offset. For each presentation, the interferer was randomly selected from a long sample (SSN: 22 s, ISTS: 52 s).

The listeners received approximately 20 min of training before the SRTs were measured. In the training session, the first sentence was presented at a very low SNR. The SNR was increased in steps of 2 dB until all five words were repeated correctly. The test subjects were allowed to guess and the recognized words were repeated verbally to the experimenter and registered without feedback. For the following sentence, the SNR was decreased by 4 dB and again increased in 2 dB steps until all the words were repeated correctly.

In the test session, a list of ten sentences was used to measure the SRT for a given run. The procedure for the presentation of the first sentence was the same as in the training session. However, for the presentation of the remaining nine sentences, the SNR followed a simple adaptive procedure: If all words were repeated correctly, the SNR was decreased by 2 dB; otherwise the SNR was increased by 2 dB. The measured SRT was the average of the last 8 SNRs from presentation number 4 to 11, where the last presentation is the SNR determined after the last sentence although there is no sentence presented. Five runs were conducted for each condition and the average of these SRTs produced the final SRT.

IV. RESULTS

The left panel of Fig. 4 shows the average SRTs obtained with the two maskers (SSN and ISTS) in the four different experimental conditions. As expected, the SRTs for the ISTS masker (gray lines) are much lower than for the SSN masker (black lines). Furthermore, the results show that, overall, the 300-Hz envelope conditions leads to lower SRTs than the 30-Hz envelope conditions and that the BB conditions leads to lower SRTs than the HP filtered conditions. A repeated measures analysis of variance (ANOVA) confirmed this by showing a main effect of masker type [$F(1,4) = 67.7$, $p < 0.005$], envelope filter conditions [$F(1,4) = 80.0$, $p < 0.001$], and HP filter conditions [$F(1,4) = 428.5$, $p < 0.0001$]. Furthermore, the ANOVA showed a significant interaction between masker type and HP filtering [$F(1,4) = 10.01$, $p < 0.05$], masker type and envelope filtering [$F(1,4) = 70.83$, $p < 0.005$], HP filtering and envelope filtering [$F(1,4) = 22.55$, $p < 0.01$], as well as masker type, HP filtering, and envelope filtering [$F(1,4) = 14.99$, $p < 0.05$]. Clearly, all of these interactions are due to the HP30 condition, which is a very important result.

A *post hoc* analysis using Tukey's HSD test, showed that for the SSN masker, there was no significant difference between the 300-Hz (solid line) and the 30-Hz (dashed line) envelope filtering in both BB [$p = 0.95$] and HP [$p = 0.53$] filtered conditions, indicating that envelope periodicity does not contribute to speech perception in stationary noise. For the ISTS masker, the SRT in the BB30 condition was about 3 dB higher than for the BB300 condition [$p < 0.0001$], while the SRT in the HP30 condition was 7 to 8 dB higher than in the HP300 condition [$p < 0.0001$]. Thus, as confirmed by the ANOVA, there is a clear interaction between envelope filtering and the reduction of resolved harmonics via HP filtering. The results for the ISTS masker indicate that, for a competing talker, envelope periodicity plays a substantial role for MR. However, the contribution of envelope periodicity is considerably smaller when the listeners can also rely on $F0$-related information from resolved harmonics.

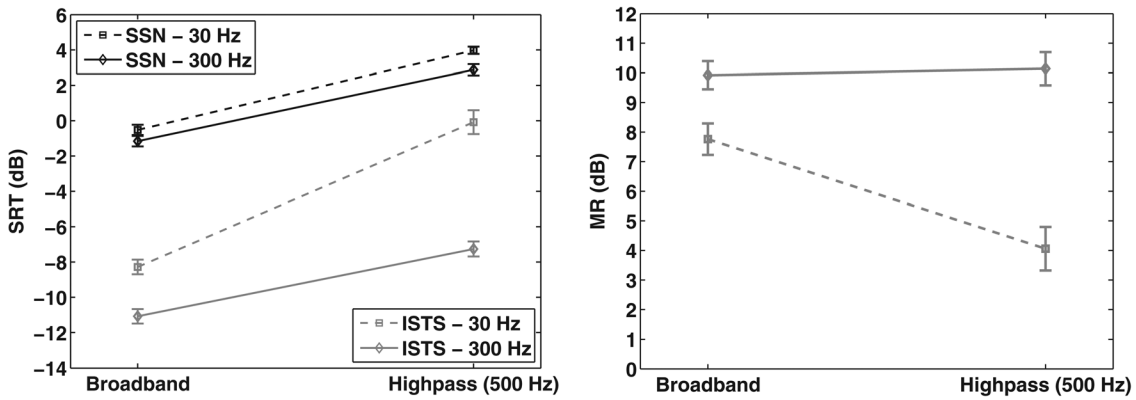

FIG. 4. The left panel shows the average SRTs obtained with the SSN masker (black lines) and the ISTS masker (gray lines) in the four different conditions. The results obtained using the 300- and 30-Hz envelope filter in the vocoder are represented by the solid and dashed lines, respectively. Whether the stimuli were broadband or HP filtered is denoted on the abscissa. The right panel shows the MR calculated as the difference in SRTs between the SSN and ISTS maskers from the left panel. Error bars indicate one standard error.

The right panel of Fig. 4 shows the MR for the ISTS in the four conditions, representing the difference in SRT between the SSN and ISTS maskers. The conditions are indicated by the same symbols and line styles as used in the left panel. A repeated measures ANOVA confirmed that the MR differed significantly across envelope filter conditions [$F(1,4) = 70.83$, $p < 0.002$], HP filter conditions [$F(1,4) = 10.01$, $p < 0.04$], and that there was a significant interaction between envelope and HP filter conditions [$F(1,4) = 14.99$, $p < 0.02$]. *Post hoc* analysis using the Tukey's HSD test indicated there was no effect of reducing the *F*0-related information from low-order resolved harmonics on the MR obtained with a competing talker [difference between BB300 and HP300 ($p = 0.99$)]. In contrast, there was a small but significant reduction of the MR (≈2 dB) when envelope periodicity was attenuated [difference between BB300 and BB30 ($p < 0.05$)]. However, when *F*0-related information from both resolved and high-order unresolved harmonics was reduced, a large reduction in the MR was observed (≈6 dB) [difference between BB300 and HP30 ($p < 0.0001$)].

V. DISCUSSION

A. Summary of the main results

The results from the present study show that HP filtering the overall stimuli at 500 Hz had no effect on MR. However, LP filtering the envelopes at 30 Hz reduced MR by approximately 2 dB. Importantly, when both processing schemes were applied, MR was strongly reduced (≈6 dB). Thus, as expected, the results suggest that *F*0 information from low-frequency harmonics contribute to MR. However, the results further suggest that envelope periodicity is also important for MR. Thus, while several studies have shown that low-order resolved harmonics dominate over high-order unresolved harmonics in pitch perception (e.g., Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Plomp, 1967; Micheyl and Oxenham, 2007), the results from the present study suggest that the envelope periodicity produced by unresolved harmonics is important for speech perception in the presence of a competing talker. However, further work is needed to clarify the importance of envelope periodicity for speech perception in other kinds of non-stationary noise and whether it is the presence of envelope periodicity in the masker, the signal, or both that is important for MR.

B. The role of resolved and unresolved harmonics for MR

The finding that the listeners were able to achieve a large MR even with reduced *F*0 information from resolved harmonics supports the results of Oxenham and Simonson (2009). It is also consistent with the indications of Stone *et al.* (2008) who used both a noise- and a tone vocoder to investigate the importance of *F*0-related envelope periodicity cues for speech perception. However, they did not explicitly investigate MR, i.e., the advantage of a competing talker relative to a stationary interferer. Qin and Oxenham (2005) found that the ability of NH listeners to recognize two concurrent vowels was significantly reduced when the stimuli were processed by a noise vocoder. They ascribed this to a poor spectral representation of the lower-order harmonics and that *F*0 information carried by the temporal envelope of higher-order harmonics may not be sufficient to support MR. This seems to be in contrast to the results of the present study. However, it could also suggest that the listeners in the present study mainly relied on sequential cues for source segregation not available in a concurrent-vowel identification task. Another explanation for the difference between the two studies could be the use of noise-band carriers in Qin and Oxenham (2005), which might have distorted the *F*0 information carried by the temporal envelope.

Oxenham and Simonson (2009) used LP filtering with a cutoff frequency of 1200 Hz and HP filtering with a cutoff frequency of 1500 Hz to isolate the effects of resolved and unresolved harmonics. In both conditions, the MR was found to be greatly reduced, indicating that resolved and unresolved harmonics separately did not contribute substantially to MR. This differs from the results of the present study where *F*0-related information from resolved harmonics and the envelope periodicity of unresolved harmonics, when considered separately, contributed substantially to MR. However, these differences in the results are likely due to differences in the procedures. Oxenham and Simonson (2009) suggested that the greatly reduced MR in both of their conditions could have been caused by a large reduction of the perceptual redundancy of the target speech due to the filtering (i.e., reduced bandwidth) and that MR might be determined mainly by the redundancy of *F*0 information. In the present study, the filtering of the envelope fluctuations did not reduce the bandwidth of the overall spectro-temporal energy pattern of the speech signal and the HP filtering used a relatively low cut-off frequency. Thus, the redundancy of speech information in the processed stimuli of the present study was probably preserved or only slightly reduced. Based on this, it seems that as long as redundancy is preserved, *F*0-related information from both resolved harmonics and envelope periodicity of unresolved harmonics contribute substantially to MR.

The results of the present study, that envelope periodicity is important for MR, appear to conflict with the results of previous work, such as studies by Bird and Darwin (1998) or Culling and Darwin (1993), which suggested that the effect of *F*0 differences is stronger at low rather than high frequencies. However, this apparent conflict stems from a fundamental difference in the paradigms used. In the studies of Bird and Darwin (1998) or Culling and Darwin (1993), listeners were presented with concurrent speech or vowel stimuli synthesized with differences in *F*0, either with consistent or inconsistent *F*0 information across frequency bands. Thus, in those studies, some *F*0 information was always present (although possibly inconsistent) across all frequency bands. In the present study, we examined the case where the *F*0 information available at low frequencies was reduced and found that MR is substantially reduced when envelope periodicity is also removed. Thus, listeners appear to make significant use of *F*0 information available from envelope periodicity, particularly when *F*0 information from resolved harmonics is reduced. When taken all together, these results suggest that in the

artificial condition where *F*0 information is not consistent across frequency bands and the masker and target share the same *F*0 in low spectral regions, the benefit from envelope periodicity in high spectral regions may be reduced.

C. Possible connections to reduced MR in HI listeners

Since the results of the present study indicate that *F*0 information plays an important role for speech perception in the presence of a competing talker, it is likely that HI listeners, who often experience great difficulties in such a condition (e.g., Festen and Plomp, 1990; Bernstein and Grant, 2009; Christiansen and Dau, 2012), might have difficulties in the processing of *F*0 information. However, we can only speculate about why these difficulties arise. Deficits in *F*0 processing could be a general deficit in extracting *F*0 information from the input stimuli, even if the stimuli only consist of a single talker in quiet conditions. Since envelope periodicity was found to be sufficient for MR in NH listeners, difficulties in the processing of *F*0 are probably related to deficits in the ability to extract *F*0 information from the temporal response of the auditory nerve fibers. This would indicate problems with phase-locking at frequencies as low as 100 to 250 Hz. However, this seems unlikely since NH listeners show robust phase-locking up to at least 1200 Hz (e.g., Santurette and Dau, 2012; Heinz, 2012) and there is some evidence that phase-locking is not reduced in the case of sensorineural hearing loss (Heinz, 2012). However, it is also possible that the extraction of *F*0 information from a single source is more or less intact, but that HI listeners have deficits in the processing of *F*0 information from simultaneous sources and thereby have difficulties separating the target speech from an interfering talker. This could be due to the interaction of source carriers within an auditory filter, which might be more pronounced in HI listeners with reduced frequency selectivity. A larger interaction of source carriers due to reduced frequency selectivity might be simulated in NH listeners in future experiments using broader filters in the vocoder presented in the current study. Further experiments with both NH and HI listeners are needed in order to clarify how hearing impairment can affect the processing of high-frequency envelope cues.

D. Implications for auditory modeling

The finding that envelope periodicity contributes substantially to MR indicates that the auditory system analyses modulation frequencies well beyond 30 Hz. If these fluctuations are indeed used to extract *F*0 information, this suggests that the analysis may extend to 300 Hz, approximately the upper limit for female speech or possibly even higher for some children's voices. Behavioral studies comparing MR across gender and age (i.e., child vs adult speech) may help clarify this upper limit. The auditory system has been shown to perform a frequency selective analysis of envelope fluctuations, which has been modeled by a modulation filterbank similar to the modeling of cochlear filters (Dau *et al.*, 1997; Ewert and Dau, 2000). A modulation-frequency specific analysis has also been found to be crucial for the prediction of speech intelligibility (Steeneken and Houtgast, 1980; Elhilali *et al.*, 2003; Jørgensen and Dau, 2011) and speech quality (Kim, 2005). However, these speech perception models only consider frequency modulations up to 32 or 64 Hz. The results from the present study indicate that, in certain conditions, auditory models should include an analysis of relatively high modulation frequencies, possibly up to several hundred hertz.

E. Stationary noise SRT and MR

It is well known that HI listeners exhibit higher SRTs in stationary noise than NH listeners. Recently, Bernstein and Grant (2009) suggested that the reduced MR exhibited by HI listeners might be due to the MR being measured with reference to a higher SRT in stationary noise (where the benefit from listening in the dips of the noise might be limited). Bernstein and Grant (2009) found a strong relation between stationary-noise SRT and MR. In a later study, Bernstein and Brungart (2011) found the MR obtained with different types of processed speech to be fully accounted for by the stationary-noise SRT. However, a recent study by Christiansen and Dau (2012) found MR to be only partly explained by the stationary-noise SRT.

In the present study, the large differences in MR observed across conditions cannot be explained by differences in SRT in stationary noise. For the BB300 and HP300 conditions, no difference in MR was observed even though the SRTs in stationary noise differ by almost 4 dB. In contrast, large differences in MR were observed across envelope filtering conditions (i.e., BB30 vs BB300 and HP30 vs HP300) even though the SRTs in stationary noise are similar. Thus, the current results support the findings of Christiansen and Dau (2012) and some of the results of Bernstein and Grant (2009) suggesting that, for competing speech, the stationary-noise SRT only partly influences MR.

VI. SUMMARY AND CONCLUSIONS

The present study investigated the contribution of *F*0-related envelope periodicity to MR obtained using a competing talker and stationary speech-shaped noise as maskers. This was done by LP filtering the envelope fluctuations using an amplitude- and frequency-modulated tone-vocoder. The contribution of *F*0 information from resolved harmonics was also investigated by removing some of these using a 500-Hz HP filter.

Envelope periodicity was found to be important for MR obtained with a competing talker. The presence of either envelope periodicity or resolved harmonics were both found to be sufficient for MR. These findings suggest that, for some situations, auditory models may require an analysis of modulation frequencies spanning the range of *F*0 produced by the talkers of the stimuli. Further work is needed to determine if the reduced MR exhibited by HI listeners is indeed due to deficits in the processing of *F*0 information.

ACKNOWLEDGMENTS

We wish to thank our colleagues at the Centre for Applied Hearing Research for valuable comments and

stimulating discussions. We are grateful to all the listeners for their participation in many hours of testing. This work has been partly supported by the Danish research council and partly by Oticon, Widex, and GN Resound through a research consortium.

Baer, T., and Moore, B. C. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Bernstein, J. G. W., and Brungart, D. S. (**2011**). "Effects of spectral smearing and temporal fine-structure distortion on the fluctuating-masker benefit for speech at a fixed signal-to-noise ratio," J. Acoust. Soc. Am. **130**, 473–488.

Bernstein, J. G. W., and Grant, K. W. (**2009**). "Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **125**, 3358–3372.

Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.

Brokx, J., and Nooteboom, S. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Christiansen, C., and Dau, T. (**2012**). "Relationship between masking release in fluctuating maskers and speech reception thresholds in stationary noise," J. Acoust. Soc. Am. **132**, 1655–1666.

Cohen, M. A., Grossberg, S., and Wyse, L. L. (**1995**). "A spectral network model of pitch perception," J. Acoust. Soc. Am. **98**, 862–879.

Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.

Darwin, C. J. (**1997**). "Auditory grouping," Trends Cogn. Sci. **1**, 327–333.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

de Cheveigné, A. (**1998**). "Cancellation model of pitch perception," J. Acoust. Soc. Am. **103**, 1261–1271.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2003**). "Recovery from prior stimulation: Masking of speech by interrupted noise for younger and older adults with normal hearing," J. Acoust. Soc. Am. **113**, 2084–2094.

Duquesnoy, A. J., and Plomp, R. (**1983**). "The effect of a hearing aid on the speech-reception threshold of hearing-impaired listeners in quiet and in noise," J. Acoust. Soc. Am. **73**, 2166–2173.

Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Commun. **41**, 331–348.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gustafsson, H. A., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.

Heinz, M. G. (**2012**). "Physiological correlates of perceptual deficits following sensorineural hearing loss," Acoust. Today **8**, 34–40.

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (**2010**). "Development and analysis of an International Speech Test Signal (ISTS)," Int. J. Audiol. **49**, 891–903.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

Jørgensen, S., and 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.

Kim, D.-S. (**2005**). "Anique: An auditory model for single-ended speech quality estimation," IEEE Trans. Speech Audio Process. **13**, 821–831.

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.

Lorenzi, C., Husson, M., Ardoint, M., and Debruille, X. (**2006**). "Speech masking release in listeners with flat hearing loss: Effects of masker fluctuation rate on identification scores and phonetic feature reception," Int. J. Audiol. **45**, 487–495.

MacDonald, E. N., Pichora-Fuller, M. K., and Schneider, B. A. (**2010**). "Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech," Hear. Res. **261**, 63–66.

Metselaar, M., Maat, B., Krijnen, P., Verschuure, H., Dreschler, W., and Feenstra, L. (**2008**). "Comparison of speech intelligibility in quiet and in noise after hearing aid fitting according to a purely prescriptive and a comparative fitting procedure," Eur. Arch. Otorhinolaryngol. **265**, 1113–1120.

Micheyl, C., and Oxenham, A. J. (**2007**). "Across-frequency pitch discrimination interference between complex tones containing resolved harmonics," J. Acoust. Soc. Am. **121**, 1621–1631.

Nguyen, D. P., Wilson, M. A., Brown, E. N., and Barbieri, R. (**2009**). "Measuring instantaneous frequency of local field potential oscillations using the Kalman smoother," J. Neurosci. Methods **184**, 365–374.

Nielsen, J. B., and Dau, T. (**2009**). "Development of a Danish speech intelligibility test," Int. J. Audiol. **48**, 729–741.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Oxenham, A. J., and Simonson, A. M. (**2009**). "Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference," J. Acoust. Soc. Am. **125**, 457–468.

Palmer, A., and Russell, I. (**1986**). "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," Hear. Res. **24**, 1–15.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient auditory filterbank based on the gammatone function," in *Proceedings of the Meeting of the IOC Speech Group on Auditory Modelling at RSRE*, December 14–15.

Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., Pass, H. E., and Brown, S. (**2007**). "Temporal jitter disrupts speech intelligibility: A simulation of auditory aging," Hear. Res. **223**, 114–121.

Plomp, R. (**1967**). "Pitch of complex tones," J. Acoust. Soc. Am. **41**, 1526–1533.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on f0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.

Qin, M. K., and Oxenham, A. J. (**2006**). "Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech," J. Acoust. Soc. Am. **119**, 2417–2426.

Santurette, S., and Dau, T. (**2012**). "Relating binaural pitch perception to the individual listener's auditory profile," J. Acoust. Soc. Am. **131**, 2968–2986.

Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.

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

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region," J. Acoust. Soc. Am. **124**, 2272–2282.

Wightman, F. L. (**1973**). "The pattern-transformation model of pitch," J. Acoust. Soc. Am. **54**, 407–416.

Xu, L., and Zheng, Y. (**2007**). "Spectral and temporal cues for phoneme recognition in noise," J. Acoust. Soc. Am. **122**, 1758–1764.