

Masking release for low- and high-pass-filtered speech in the presence of noise and single-talker interference

Andrew J. Oxenham^{a)} and Andrea M. Simonson

Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, Minnesota 55455

(Received 18 October 2007; revised 24 September 2008; accepted 13 October 2008)

Speech intelligibility was measured for sentences presented in spectrally matched steady noise, single-talker interference, or speech-modulated noise. The stimuli were unfiltered or were low-pass (LP) (1200 Hz cutoff) or high-pass (HP) (1500 Hz cutoff) filtered. The cutoff frequencies were selected to produce equal performance in both LP and HP conditions in steady noise and to limit access to the temporal fine structure of resolved harmonics in the HP conditions. Masking release, or the improvement in performance between the steady noise and single-talker interference, was substantial with no filtering. Under LP and HP filtering, masking release was roughly equal but was much less than in unfiltered conditions. When the average F_0 of the interferer was shifted lower than that of the target, similar increases in masking release were observed under LP and HP filtering. Similar LP and HP results were also obtained for the speech-modulated-noise masker. The findings are not consistent with the idea that pitch conveyed by the temporal fine structure of low-order harmonics plays a crucial role in masking release. Instead, any reduction in speech redundancy, or manipulation that increases the target-to-masker ratio necessary for intelligibility to beyond around 0 dB, may result in reduced masking release. © 2009 Acoustical Society of America.

[DOI: 10.1121/1.3021299]

PACS number(s): 43.71.Gv, 43.66.Hg [RLF]

Pages: 457–468

I. INTRODUCTION

A common complaint of hearing-impaired listeners and cochlear-implant users relates to their difficulty in understanding speech in complex acoustic backgrounds. One illustration of this problem comes from studies that compare speech intelligibility in steady noise with intelligibility in spectrally and/or temporally fluctuating maskers. Speech intelligibility, as measured in normal-hearing listeners, improves dramatically when temporal fluctuations are introduced into a noise masker, or when the noise masker is replaced by a single-talker interferer with the same long-term power spectrum (e.g., Miller and Licklider, 1950; Festen and Plomp, 1990). This “masking release” has been ascribed to normal-hearing listeners’ ability to “listen in the valleys,” making use of favorable target-to-masker ratios (TMRs) in brief (or narrowband) low-energy temporal (or spectral) epochs in the masker. In contrast, hearing-impaired listeners typically show much less release from masking when the masker is changed from steady noise to a fluctuating noise or single-talker interferer (e.g., Festen and Plomp, 1990; Peters *et al.*, 1998), and cochlear-implant users have been found to show either no masking release, or even an increase in masking, under similar conditions (Nelson *et al.*, 2003; Stickney *et al.*, 2004). For hearing-impaired listeners, the lack of masking release has traditionally been explained in terms of a putative loss of temporal and spectral resolution, limiting their ability to make use of brief or narrowband dips in

masker energy. Similarly, for cochlear-implant users, at least some of the deficit may be due to severely reduced spectral resolution.

More recently, a number of studies have suggested that deficits in fundamental frequency (F_0) coding and pitch perception may also be a possible cause for the lack of masking release in listeners with hearing impairment (Summers and Leek, 1998) or cochlear implants (Nelson and Jin, 2004; Stickney *et al.*, 2004). In contrast to explanations based on a loss of spectral or temporal resolution, a reduction in masking release due to reduced pitch salience does not necessarily imply that the target was masked by the interferer in the traditional sense. Instead, it suggests that the cues necessary to distinguish the target from the interferer were not sufficiently salient for successful perceptual segregation to occur. The failure to distinguish and segregate a target from an interferer could result in the masker fluctuations being confused for those of the target, thus actually producing more (rather than less) masking. This phenomenon has been referred to as a form of “informational masking,” rather than the more traditional “energetic masking” that is believed to dominate for a steady-noise masker (e.g., Brungart, 2001). In cases where the interferer is speech, temporal fine-structure cues may help to distinguish between the different F_0 s, or F_0 trajectories, of the target and the interferer (Brox and Nooteboom, 1982; Bird and Darwin, 1998; Assmann, 1999; Stickney *et al.*, 2007). In cases where the interferer is a modulated noise, the fact that parts of the speech are voiced may be sufficient to produce some segregation between the speech and the noise (Qin and Oxenham, 2003; Lorenzi *et al.*, 2006).

In normal hearing, pitch information appears to be conveyed primarily by the temporal fine structure of the lower-

^{a)}Author to whom correspondence should be addressed. Electronic mail: oxenham@umn.edu

order harmonics, which are thought to be resolved in the peripheral auditory system. The ability of listeners to discriminate small differences in F_0 is considerably better for low-order harmonics than high-order (unresolved) harmonics (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Bernstein and Oxenham, 2003), and when F_0 information from low-order harmonics conflicts with that from high-order harmonics, the low-order harmonics usually dominate the percept, both for isolated harmonic complex tones (Plomp, 1967; Micheyl and Oxenham, 2007) and for speech (Bird and Darwin, 1998). In fact, studies using harmonic tone complexes consisting only of high-order unresolved harmonics suggest that listeners are not capable of extracting the pitches of two complexes presented simultaneously in the same spectral region (Carlyon, 1996; Micheyl *et al.*, 2006).

Hearing-impaired listeners typically have less access to resolved harmonics, due in part to abnormally broad auditory filters and, in many cases, poorer coding of single pure tones (e.g., Bernstein and Oxenham, 2006). It has also been claimed that hearing-impaired listeners may have a reduced ability to process temporal fine-structure information (Lorenzi *et al.*, 2006), although it is difficult to separate this from the effects of poorer frequency selectivity. Whatever the underlying mechanisms, Lorenzi *et al.* (2006) showed that poorer temporal fine-structure processing seems to correlate negatively with masking release when comparing speech reception in steady noise with that in fluctuating noise.

The coding schemes used in cochlear implants do not typically convey temporal fine-structure information in the waveforms of each electrode, and the spectral representation is too coarse to provide temporal fine structure in terms of place of excitation of individual harmonics. Thus, if the accuracy of pitch coding plays a major role in masking release, one reason why both hearing-impaired listeners and cochlear-implant users show severely reduced amounts of masking release may be the reduced pitch salience experienced by them, which is induced by reduced or no access to the temporal fine structure associated with low-order resolved harmonics.

Given the proposed link between pitch salience and temporal fine structure on the one hand, and masking release on the other, an obvious prediction is that normal-hearing listeners should show reduced masking release if pitch salience and temporal fine-structure cues are disrupted in some way. Recent studies using acoustic simulations of cochlear-implant processing in normal-hearing listeners seem to support this prediction (e.g., Nelson *et al.*, 2003; Qin and Oxenham, 2003, 2006; Gnansia *et al.*, 2008; Hopkins *et al.*, 2008). For instance, Qin and Oxenham (2003) measured speech intelligibility in various types of interferers with the stimuli processed through noise-excited envelope vocoders that reduce spectral resolution (depending on the number of frequency channels) and replace the original temporal fine structure with a noise carrier. They found no release from masking (and even a slight increase in masking) when a steady noise was replaced by a single talker (male or female) interferer, even with the relatively good spectral resolution afforded by 24-channel simulations. Less masking release

was also observed for a noise masker that was modulated with the temporal envelope of speech. Nevertheless, vocoder simulations distort the signal in other ways than simply reducing access to the original temporal fine structure; they also reduce spectral resolution somewhat (even in the 24-channel case) and, for noise-excited vocoders, introduce spurious envelope fluctuations that may also affect performance (Whitmal *et al.*, 2007). More recent studies (Gnansia *et al.*, 2008; Hopkins *et al.*, 2008) have used even more channels (32 in the case of Gnansia *et al.*, 2008) and tonal carriers, and have still found reduced masking release. However, even here, it is difficult to rule out the role of potential temporal envelope distortions due, for instance, to beating between neighboring carriers.

The present study provides a more direct test of the hypothesis that the pitch information provided by the temporal fine structure of low-order resolved harmonics is important for masking release. Sentences were presented in steady or fluctuating backgrounds in unfiltered, low-pass-(LP), or high-pass-(HP) filtered conditions. The rationale was that HP filtering the stimuli should eliminate the low-order harmonics necessary for accurate pitch perception, while LP filtering the stimuli should retain these low-order harmonics. Thus, if low-order resolved harmonics are necessary for masking release, then masking release should be observed in the LP-filtered condition, while little or no masking release should be observed in the HP-filtered condition. Experiment 1 tests this hypothesis by comparing speech intelligibility in steady noise with that in the presence of a single-talker interferer in broadband (BB), LP-filtered, and HP-filtered conditions. Experiment 2 increases the pitch differences between the target and interfering speech by artificially shifting the pitch contour of the interferer to lower F_0 s. Experiment 3 tests whether temporal fine structure is important for situations in which the masker is a temporally fluctuating noise, as suggested by some previous studies (Qin and Oxenham, 2003; Lorenzi *et al.*, 2006).

II. EXPERIMENT 1: INTELLIGIBILITY IN STEADY NOISE AND SINGLE-TALKER MASKERS

A. Methods

1. Stimuli

The target sentences were Hearing in Noise Test (HINT) sentence lists (Nilsson *et al.*, 1994), spoken by a male talker. The average F_0 of these HINT sentences is approximately 110 Hz, with an average within-sentence standard deviation of 24 Hz. Speech-shaped noise maskers consisted of BB Gaussian noise that was filtered to match the long-term power spectrum of the HINT sentences. Competing-talker maskers were created using recordings of a subset of IEEE sentences (IEEE, 1969), also produced by a male. These sentences were also spectrally shaped to match the long-term power spectrum of the HINT sentences. The sentences were concatenated and then broken into 4-s blocks to be used as maskers, without regard to the beginnings of sentences. In other words, a 4-s block could start in the middle of a sentence. The gaps that occurred between concatenated sentences occurred anywhere within each 4-s block, and had an

average duration of 135 ms (with gaps defined as contiguous periods during which the level remained at least 35 dB below the long-term rms level of the speech). The average F_0 of the competing-talker masker was approximately 106 Hz, with an average within-sentence standard deviation of 28 Hz.

The target sentences and maskers were added together at the appropriate TMR for a given condition before further filtering. The filtered stimuli were created by either LP or HP filtering the combined target and masker using a fourth-order Butterworth filter. The LP condition involved stimuli that were LP filtered with a cutoff frequency of 1200 Hz; the HP condition used a cutoff frequency of 1500 Hz. These particular cutoff frequencies were chosen in an attempt to equate overall performance in the speech-shaped steady-noise conditions, based on the results from a pilot study. A complementary off-frequency noise was subsequently added to the HP- and LP-filtered stimuli. The noise had the same long-term spectrum (before filtering) as the HINT target speech, and its unfiltered level was set to be 12 dB below the level of the unfiltered target speech. In the LP condition, the complementary noise was HP filtered at 1200 Hz; in the HP condition, the complementary noise was LP filtered at 1500 Hz (also using fourth-order Butterworth filters). The purpose of the complementary noise was to limit the possibility for “off-frequency listening” to information in the slopes of the filters, and to mask any low-frequency distortion products that might be generated by the speech in the HP-filtered conditions.

In all trials, the speech or noise masker and the complementary noise (when present) had a total duration of 4 s, gated on and off with 5-ms ramps. The target sentences ranged in duration from 1.17 to 2.46 s, and began 750 ms after the beginning of the masker.

2. Subjects

Twenty-four subjects (18 female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing, and had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz. Their ages ranged from 19 to 55, with a mean age of 25.3. Subjects were paid for their participation.

3. Procedure

Three groups of eight subjects each were tested using one of the three filter conditions (unfiltered, HP filtered, and LP filtered). Subjects were told that they would hear distorted sentences in a noisy background and they were instructed to type what they heard via a computer keyboard. It was explained that some of the utterances would be hard to understand, and that they should make their best guess at as many words as possible.

Approximately 1 h of practice was provided before testing. Practice target sentences were taken from 12 lists of sentences from the IEEE/TIMIT corpus (IEEE, 1969). These sentences were produced by a different talker from that used to create the maskers. Subjects heard half the sentences in a background of speech-shaped noise, and the other half in a

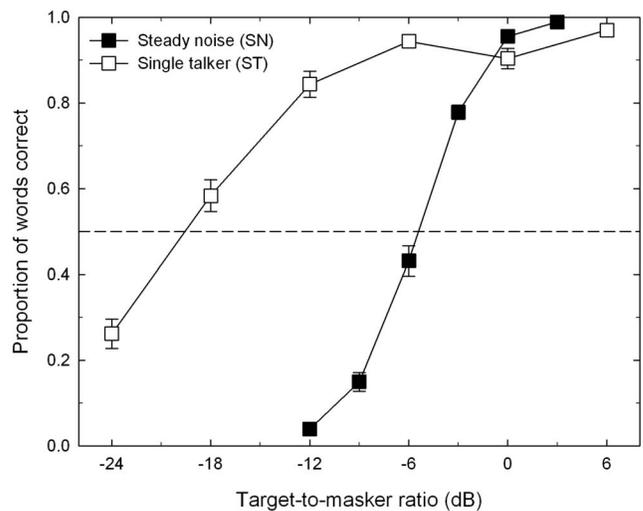


FIG. 1. Proportion of words correctly reported from HINT sentence lists as a function of TMR. Open symbols represent performance in the presence of a single-talker interferer and filled symbols represent performance in the presence of a steady-noise interferer. The square symbols in this and all other figures represent unfiltered conditions. Error bars represent ± 1 standard error of the mean across subjects.

competing-talker background, at several TMRs in BB, LP, and HP conditions. After each practice sentence, listeners were prompted to type what they had heard via a computer keyboard. Feedback was then provided by displaying the sentence text on the computer monitor. Listeners had the option to repeat the sentence as often as they wished before moving on to the next sentence.

In the actual experiment, the stimuli were presented at six different TMRs for each masker type (speech-shaped noise, competing talker), resulting in 12 conditions. Two lists, each comprising ten HINT target sentences, were presented for each condition, resulting in a total of 24 lists of HINT sentences in the experiment. No target sentence was presented more than once to any subject. Each condition was presented once before any was repeated. The order of presentation was counterbalanced across subjects for the masker type, and always progressed from highest (easiest) to lowest (hardest) TMR. No feedback was given during the actual experiment, which took approximately 1 h and immediately followed the practice session.

For both the practice and the actual experiment, the stimuli were presented diotically at an overall level of 70 dB SPL (after processing) in each ear over Sennheiser HD580 headphones. The stimuli were played out via a Lynx22 (LynxStudio) sound card at 16 bit resolution with a sampling rate of 22.05 kHz. After the session, the listeners’ responses were scored offline. All words except for “a” and “the” were counted, and the percentage of correct words in each condition was calculated. Obvious misspellings were counted as correct.

B. Results and discussion

The mean results from the unfiltered conditions are shown in Fig. 1, with proportion of words correctly reported plotted as a function of TMR. Error bars represent ± 1 standard error across the eight subjects. The raw percent-correct

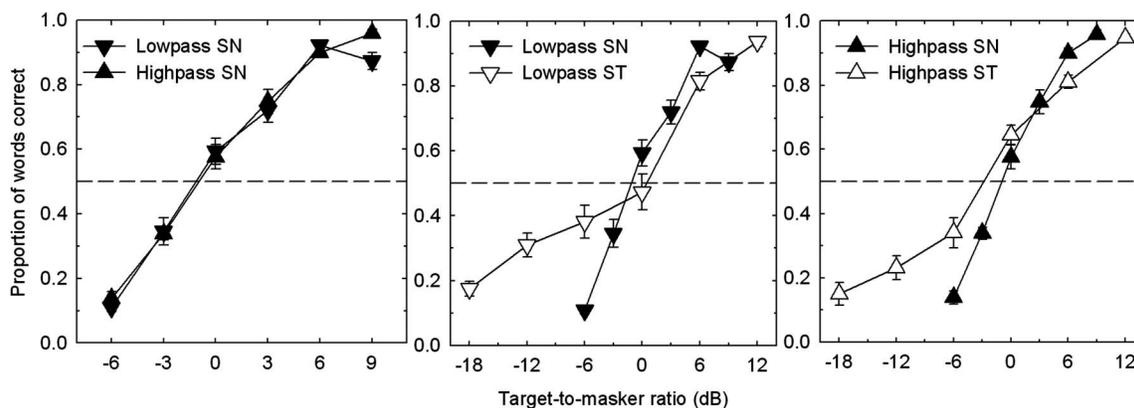


FIG. 2. Proportion of words correctly reported from HINT sentence lists as a function of target-to-masker ratio. Upward- and downward-pointing triangles represent HP and LP conditions, respectively. The left panel shows results in steady noise (filled symbols); the center and right panels compare those results with performance in a single-talker interferer (open symbols) in LP and HP conditions, respectively. Error bars represent ± 1 standard error of the mean across subjects.

scores are shown; however, for the purposes of statistical analyses, the data were subjected to an arcsine transform. Filled symbols denote performance in steady-noise and open symbols denote performance in the single-talker background. As has been shown in many previous studies (e.g., Duquesnoy, 1983; Festen and Plomp, 1990; Peters *et al.*, 1998; Summers and Molis, 2004), a single-talker interferer produces considerably less masking than a steady noise with the same overall level and long-term spectrum. The amount of masking release can be quantified in various ways. One way is to consider the increase in proportion of words correctly reported at a given TMR (Takahashi and Bacon, 1992; Jin and Nelson, 2006). Quantified in this way, masking release was about 80 percentage points at a TMR of -12 dB, 50 percentage points at a TMR of -6 dB, and no masking release (or a “negative release” of 5 percentage points) at 0 dB TMR, perhaps because of ceiling effects. Another way to quantify masking release is to determine the change in TMR necessary to obtain 50% of words correct; this TMR is known as the speech reception threshold (SRT) (e.g., Qin and Oxenham, 2003).¹ The SRTs were estimated from our data by fitting a logistic function to the individual data from all TMRs, and determining the TMR at which the fitted functions crossed the 50% line. The logistic functions were fitted using maximum likelihood with an assumed binomial distribution. The SRTs, averaged across the SRTs from logistic fits to the data from individual subjects, were -5.6 and -19.5 dB for the steady-state noise and single-talker interferer, respectively, resulting in an average masking release of around 14 dB.

The left panel of Fig. 2 shows the results from the LP and HP-filtered conditions in a steady-noise background. The similarity of the two curves indicates that we were successful in selecting LP and HP filter cutoffs that produced roughly equal performance in steady noise for the two conditions. A mixed-model analysis of variance (ANOVA) on the arcsine transformed data confirmed a significant effect of TMR ($F_{5,70}=241.0$, $p<0.0001$) but no effect of filter type ($F_{1,14}=1.39$, $p=0.26$).² The interaction between filter type and TMR was significant ($F_{5,70}=3.84$, $p=0.01$), presumably due to the downturn in performance at the highest TMR in the LP

condition. The middle panel of Fig. 2 replots the LP results from the left panel and compares them to the LP results using a competing-talker interferer. Some masking release is present, at least at negative TMRs, but it is considerably less than that found in the unfiltered conditions. The right panel shows the same comparison of steady noise and single-talker interferer, but for the HP-filtered conditions. Again, the amount of masking release is greatly reduced, relative to the unfiltered conditions, but is similar to that found in the LP filtered conditions. The similarity of the HP and LP-filtered conditions was confirmed by a mixed-model ANOVA comparing the HP- and LP-filtered single-talker conditions, with TMR as a within-subject variable and filter condition as a between-subject variable. As expected, there was a significant effect of TMR ($F_{5,70}=263.8$, $p<0.0001$) but no main effect of filter type ($F_{1,14}<1$, n.s.).

Overall, the amount of masking release was small in both filtered conditions. In the LP condition, a within-subject ANOVA, using only TMRs that were measured in both the single-talker and steady-noise maskers, revealed no main effect of masker type ($F_{1,7}<1$, n.s.), but a significant interaction between masker type and TMR ($F_{2,14}=39.14$, $p<0.0001$). In the HP condition, the main effect of masker type failed to reach significance ($F_{1,7}=3.97$, $p=0.086$), but the interaction between TMR and masker type was significant ($F_{2,14}=13.51$, $p=0.0006$). In both HP and LP conditions, it appears that masking release is present only at negative TMRs, and is effectively absent, or even reversed, at TMRs of 0 and +6 dB.

The SRTs for the LP and HP conditions were obtained by averaging the SRTs from logistic fits to the individual data. In the LP condition, the SRT was -0.5 dB in the steady noise and -3.7 dB in the single-talker masker, resulting in an overall masking release of only 3.2 dB. The SRT for the LP single-talker masker, as estimated using a logistic function does not match the “visual” SRT well (estimated by where the interpolated data points in Fig. 2 cross the 50% line), mainly because the logistic function cannot capture the apparent dip in performance at a TMR of 0 dB. This dip was not found in the HP-filtered conditions, where the estimated SRTs for the steady noise (-0.7 dB) and the single talker

(−3.9 dB) agree better with visual inspection of the data in Fig. 2, to again produce an estimated masking release of 3.2 dB.

In summary, masking release was similar in both the LP and HP conditions and was reduced relative to the unfiltered conditions. The results suggest that masking release in normal-hearing listeners can be severely reduced simply by limiting the available frequency spectrum, without any further degradation. The results are not consistent with our initial prediction that the LP-filtered condition should result in greater masking release than that observed in the HP-filtered conditions, because of the presence of low-order resolved harmonics in the LP, but not in the HP, condition.

There are a number of possible reasons for the failure of the results to match predictions. One possible explanation is that the HP filter cutoff was not set high enough, and that some resolved harmonics were present in that condition. To test for this, we carried out an F_0 discrimination experiment using complex tones with the same long-term spectral envelope as that of the target speech. As described in the Appendix, we found very pronounced differences between the LP- and HP-filtered conditions, with large (poor) F_0 difference limens (DLs) and strong phase dependencies in the HP conditions, as expected from unresolved harmonics, and small (good) F_0 DLs and no phase dependence in the LP conditions, as expected from resolved harmonics. Thus, it seems that the filters used in this experiment were successful at restricting the stimuli in the HP condition to only unresolved harmonics.

Another possible explanation is that the average F_0 s of the target (110 Hz) and interferer (106 Hz) were very similar in this experiment. It is possible that the pitch differences were not sufficient to provide a strong segregation cue, and that other cues that help us to identify different talkers, such as vocal tract length (Darwin *et al.*, 2003), were sufficient to provide whatever small masking release was observed in both LP- and HP-filtered conditions. In fact, the similarity in average F_0 between the target and masker may explain the apparent dip in performance at 0 dB TMR in the LP condition: it may be that confusion between the target and masker was greatest when there were no overall level differences between the two (e.g., Brungart, 2001) and that other cues, such as vocal tract length, were conveyed better in the HP than in the LP condition. This idea is tested in experiment 2 by artificially lowering the F_0 of the interferer, while retaining all other speaker characteristics, such as vocal tract size, in order to increase the pitch difference between the target and interferer.

III. EXPERIMENT 2: INTELLIGIBILITY IN A SINGLE-TALKER MASKER WITH SHIFTED F_0

A. Methods

1. Subjects

Nine subjects (four female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing and had audiometric thresholds of 20 dB HL or less at octave frequencies between

250 and 8000 Hz. Their ages ranged from 19 to 31, with a mean age of 21.8. Subjects were either paid or given course credit for their participation.

2. Stimuli and procedure

The stimuli were very similar to those used for the single-talker condition in experiment 1, with the exception that the F_0 of the masker was lowered by four semitones using PRAAT software (Boersma and Weenink, 2008). The processed- F_0 interferer had an average F_0 of 85 Hz, with an average within-sentence standard deviation of 22 Hz, and was thus quite far removed from the average F_0 of the target (110 Hz). As before, the interfering speech was spectrally shaped to match the long-term spectrum of the target speech. Subjects heard all the sentences in a single-talker background, at four different TMRs, and each subject was tested using all three filter conditions (unfiltered, HP filtered, and LP filtered). The order of presentation was counterbalanced across subjects for the filter type and always progressed from highest (easiest) to lowest (hardest) TMR. As in experiment 1, the LP-filter cutoff was 1200 Hz, the HP-filter cutoff was 1500 Hz, and the complementary off-frequency noise was presented in each of the filtered conditions. Because the interferer F_0 was lower than in experiment 1, the harmonics of the interferer should have been even less resolved than those of the interferer in experiment 1. As in experiment 1, subjects were provided with approximately 1 h of training with feedback (on a different sentence corpus) before testing.

B. Results and discussion

The results are shown in Fig. 3, with the data from the steady-noise masking conditions from experiment 1 replotted as small symbols for comparison. The left panel shows results from the unfiltered condition, and the right panel shows the results from the filtered conditions. In the unfiltered condition, the shift in the interferer F_0 resulted in about the same amount of masking release as was found in the original unfiltered conditions, with a mean estimated SRT of −18.5 dB (compared with −19.5 dB in the unprocessed condition). In contrast, for both the LP- and HP-filtered stimuli, more masking release was observed with the shifted- F_0 maskers than with the original single-talker maskers. A comparison of the shifted and original single-talker conditions indicated a main effect of F_0 shift, with improved performance using the shifted- F_0 masker [main effect of shift: $F_{1,15}=4.70$, $p=0.047$ (LP) and $F_{1,15}=6.08$, $p=0.03$ (HP)]. As found in experiment 1, LP and HP filtering resulted in very similar performance overall. A within-subject ANOVA comparing LP- and HP-filtered conditions with the shifted- F_0 interferer confirmed no significant difference between the two ($F_{1,8} < 1$, n.s.), with no significant interaction between filter condition and TMR ($F_{3,24}=1.809$; $p=0.173$). Analysis in terms of SRT yielded similar conclusions. The SRTs, estimated from the logistic fits to the individual data, were −7.1 and −6.9 dB (resulting in a release from masking of about 6.6 and 6.2 dB) for the LP and HP conditions, respectively. Interestingly, in this case there was no apparent difference in performance between the LP and HP conditions

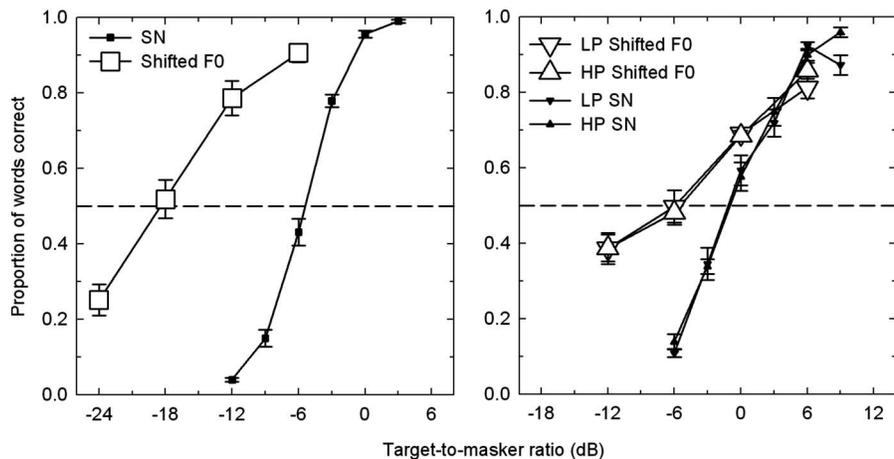


FIG. 3. Proportion of words correctly reported from HINT sentence lists as a function of TMR in conditions where the F_0 of the single-talker interferer was shifted down by four semitones. Results in steady noise from Figs. 1 and 2 are replotted as small symbols for comparison. The left panel shows results from unfiltered conditions, and the right panel shows results from the LP- and HP-filtered conditions. Error bars represent ± 1 standard error of the mean across subjects.

at a TMR of 0 dB, as was observed in experiment 1. It may be that the confusion in experiment 1 between the target and masker was greater in the LP than in the HP condition, and that a greater difference in average F_0 between the target and masker reduced the confusion, and hence the difference between the LP and HP conditions at a 0 dB TMR.

In summary, shifting the F_0 of the interfering talker increased the amount of masking release in both LP- and HP-filtered conditions. However, the overall amount of masking release was still very similar in both conditions. As in experiment 1, the results are not in line with our prediction that the more salient and accurate pitch conveyed by the lower-order resolved harmonics should lead to greater masking release in the LP- than in the HP-filtered conditions.

There are again a number of potential reasons why our expectations were not met. For instance, it may be a coincidence that the LP and HP conditions in experiments 1 and 2 both produce very similar amounts of masking release. It should be remembered that the low and high spectral regions can convey very different forms of speech information, even if the overall intelligibility is the same (e.g., Northern and Downs, 2002). In fact, both the LP and HP conditions lack a considerable portion of the spectrum between 300 and 3500 Hz, which is considered to contain the most important speech information (French and Steinberg, 1947; Assmann and Summerfield, 2004). Also, even at an acoustic level, the strength of masker fluctuations may vary somewhat in different frequency regions; for instance, Festen and Plomp (1990) found that there was typically more speech modulation energy in high-frequency regions than in low-frequency regions. Thus, it is possible that the greater amount of pitch information in the low-frequency region may be offset to some degree by the greater masker modulation depth (and hence greater scope for masking release) in the high-frequency region. Nevertheless, it seems clear that the temporal fine-structure information from low-order harmonics in voiced speech does not play a dominant role in determining masking release, at least when comparing steady-noise maskers to single-talker interferers. The aim of the final experiment was to generalize our findings to another condition for which masking release is often measured, involving masking by amplitude-modulated noise.

IV. EXPERIMENT 3: MASKING RELEASE IN SPEECH-MODULATED NOISE

A. Rationale

Hearing impairment, cochlear implants, and simulations of cochlear implants using channel vocoders all lead to reduced masking release, relative to normal hearing, when comparing modulated noise with steady-state noise maskers (e.g., Festen and Plomp, 1990; Bacon *et al.*, 1998; Peters *et al.*, 1998; Qin and Oxenham, 2003; Nelson and Jin, 2004; Jin and Nelson, 2006; Rhebergen *et al.*, 2006). Various types of modulated noise have been used in previous studies, including sinusoidally amplitude (or intensity) modulated noise, square-wave gated noise, and noise that is modulated by the broadband temporal envelope of speech from a single talker. As with masking release using single-talker interferers, it has been proposed that the reduced ability of hearing-impaired listeners and cochlear-implant users to take advantage of temporal dips in a modulated noise may be related to reduced access to temporal fine-structure cues (e.g., Qin and Oxenham, 2003; Lorenzi *et al.*, 2006). The rationale is that severely reduced (or absent) temporal fine-structure cues may render it more difficult to distinguish the fluctuations in the (partially periodic) target speech from those of the noise masker.

In this experiment, we measured speech intelligibility in the presence of speech-modulated noise. Because the temporal envelope of BB speech was used, the modulation energy was the same in both low- and high-frequency regions. This enabled us to test whether any differences emerge between masking release in the LP and HP conditions when the modulation energy of the masker is the same in both regions.

B. Methods

1. Subjects

Nine subjects (eight female) participated in this experiment. None had previously listened to HINT sentences. The subjects all reported normal hearing, and had audiometric thresholds of 20 dB HL³ or less at octave frequencies between 250 and 8000 Hz. Their ages ranged from 19 to 31, with a mean age of 20.7. Subjects were either paid or given course credit for their participation.

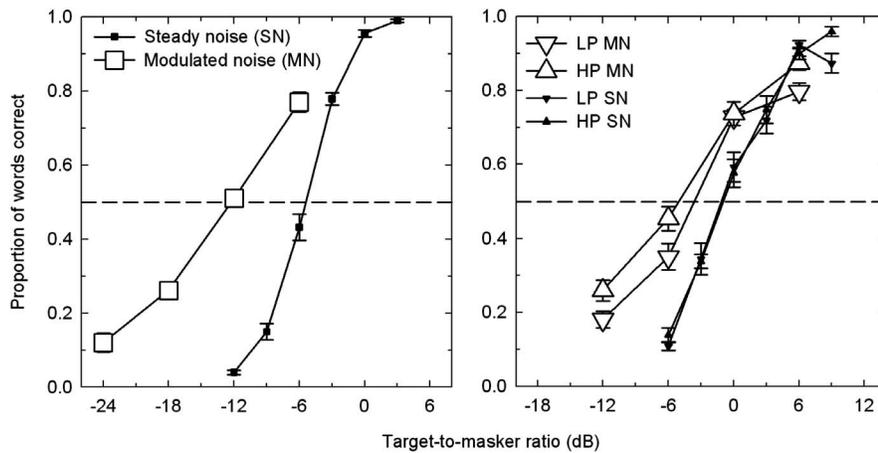


FIG. 4. Proportion of words correctly reported from HINT sentence lists as a function of TMR. Open symbols represent results in the presence of a speech-modulated noise masker. Results in steady noise from Figs. 1 and 2 are replotted as small symbols for comparison. The left panel shows results from unfiltered conditions, whereas the right panel show results from the LP- and HP-filtered conditions. Error bars represent ± 1 standard error of the mean across subjects.

2. Stimuli and procedure

Modulated-noise maskers were generated based on the envelopes of the maskers from experiment 1. Envelopes were obtained by half-wave rectifying and LP filtering the BB speech stimulus using a sixth-order Butterworth filter with a cutoff frequency of 50 Hz. The resulting envelopes were used to modulate a noise that was spectrally shaped to have the same long-term power spectrum as the target HINT sentences.

Subjects heard the target HINT sentences in a modulated-noise background at four different TMRs. As in the previous experiment, each subject was tested under all three filter conditions (unfiltered, HP filtered, and LP filtered) using a LP-filter cutoff of 1200 Hz and a HP-filter cutoff of 1500 Hz, with complementary off-frequency noise in the filtered conditions. The testing session was preceded by a 1 h training session with feedback, as described in experiment 1.

C. Results and discussion

The results are shown in Fig. 4, with the steady-noise data from experiment 1 replotted as small symbols for comparison. The left panel shows results from the unfiltered condition and the right panel shows results from the filtered conditions. The unfiltered results showed a significant amount of masking release, when compared with the speech-shaped noise condition from experiment 1, at the TMRs that both conditions had in common ($F_{1,15}=243.53$; $p<0.0001$), although performance was lower overall than that observed with the single-talker interferer in experiment 1 ($F_{1,15}=95.93$; $p<0.0001$). The finding of less masking release for speech-modulated noise than for a single-talker interferer is consistent with earlier studies (e.g., Peters *et al.*, 1998).

Both the LP and HP conditions showed a significant amount of masking release when compared with the comparable steady-noise conditions from experiment 1 (LP: $F_{1,15}=5.93$, $p=0.028$; HP: $F_{1,15}=16.46$, $p=0.001$). A within-subject comparison of the low-pass and high-pass conditions revealed that, contrary to our initial hypothesis, performance in the HP condition actually exceeded that in the LP condition overall ($F_{1,8}=33.17$, $p<0.001$), although the difference did not exceed ten percentage points at any TMR. There was no interaction between TMR and filter condition ($F_{3,24}$

$=1.10$, $p=0.37$). The mean estimated SRTs were -3.4 and -5.5 dB in the low-pass and high-pass conditions, respectively.

The fact that the LP condition did not result in more masking release than the HP condition when the masker modulation energy was equated across the two regions does not provide support for the idea that the higher masker modulation energy in the HP conditions of experiments 1 and 2 counteracted any benefit of temporal fine structure in the LP conditions. However, the results do not provide strong evidence *against* the hypothesis either. This is because there are many differences between the single-talker interferer of experiments 1 and 2 and the modulated noise of experiment 3, which make a direct comparison problematic. On one hand, the acoustic differences between the modulated noise and the target are much greater than those between the interfering talker and the target. This might imply that the pitch cues encoded in the temporal fine structure may not be as necessary when segregating speech from noise, as compared to segregating speech from speech. In other words, the more similar acoustics, as well as the intelligibility (Rhebergen *et al.*, 2005), of a speech masker may produce more informational masking than a modulated-noise masker, which may in turn be alleviated more by pitch cues. On the other hand, a speech interferer has spectral, as well as temporal, dips, which may enhance the potential amount of masking release. It is not clear which of these (or other) competing factors governs the similarities and differences between the different masker types. Nevertheless, it is clear that our initial findings of no specific benefit of pitch-related temporal fine-structure cues in the LP region generalize from single-talker interferers to temporally modulated noise maskers.

V. GENERAL DISCUSSION

A. Summary of results

Three experiments were carried out to compare word intelligibility in sentences using steady or fluctuating maskers under different spectral conditions (unfiltered, LP filtered, and HP filtered). The cutoff frequencies (LP 1200 Hz; HP 1500 Hz) were selected to produce roughly equal speech intelligibility for the two filtered conditions in the presence of steady noise and to minimize the availability

of peripherally resolved harmonics in the HP condition (see Appendix). Experiment 1 found a substantial release from masking in the unfiltered conditions when the steady noise was replaced by a single-talker interferer. The amount of masking release was greatly reduced in both the LP and HP conditions, and no significant differences between the two filtered conditions were found. Experiment 2 used the same single-talker masker as was used in experiment 1, but artificially lowered the F_0 of the masker to increase the average F_0 difference between the target and the masker. Increasing the average F_0 difference between the target and masker did not affect performance in the unfiltered conditions, where the masking release remained substantial. The increase in F_0 difference did increase the release from masking found in the two filtered conditions, but it did so by the same amount for both the LP and HP conditions. Experiment 3 measured intelligibility in the presence of a noise masker that was temporally modulated with a broadband speech stimulus. The same speech modulator was used in the unfiltered and in the LP and HP conditions. As in the previous two experiments, more masking release was observed (relative to performance in unmodulated noise) in the unfiltered conditions than in the LP and HP conditions, which again both produced similar amounts of masking release; in fact, the amount of masking release was somewhat greater for the HP than for the LP condition.

B. Possible interpretations and implications

The results are not consistent with the idea that the pitch information carried in the temporal fine structure of low-order resolved harmonics plays a crucial role in masking release. The results do not therefore seem to support earlier conjectures that loss of temporal fine-structure and pitch information may underlie the reduced masking release experienced by hearing-impaired listeners (e.g., Lorenzi *et al.*, 2006) and cochlear-implant users (e.g., Qin and Oxenham, 2003; Stickney *et al.*, 2004).

One possible interpretation is that any form of band limiting of the stimuli may result not only in poorer performance overall but also in a reduction in the advantage that listeners can take of spectrotemporal fluctuations in a masking stimulus. It may be that a large proportion of the masking release observed in BB conditions stems from the inherent redundancy in the speech signal, and that a reduction in that redundancy, by band-limiting the stimuli, reduces the scope for masking release. Loss of redundancy may also explain in part why hearing-impaired listeners, as well as simulated and real cochlear-implant users, typically exhibit less (or no) masking release. It may be, for instance, that impaired frequency resolution is sufficient to reduce redundancy to a point where little or no masking release is observed.

Another, not mutually exclusive, possibility is that masking release can only occur at relatively low TMRs. It is noteworthy that in all our conditions (filtered and unfiltered), masking release was only observed at TMRs of 0 dB or less. In the unfiltered conditions, the interpretation is made difficult by the fact that performance was near ceiling at 0 dB TMR. However, even in these conditions there is some hint

of a “crossover” around 0 dB TMR, as is more clearly observed in the filtered conditions, where the single-talker interferer actually becomes a more effective masker than the steady noise at high TMRs. As suggested by Kwon and Turner (2001), masker fluctuations can in principle hinder as well as help performance, by introducing spurious fluctuations into the target. In cases where the TMR is positive, where there is little scope for confusion between the target and the masker, it may be that the detrimental effects of spurious masker fluctuations outweigh any benefit of dips in the masker energy. A potentially related finding was recently reported by Freyman *et al.* (2008) in the domain of spatially induced (as opposed to F_0 -induced) masking release. Using noise-excited envelope-vocoded stimuli, they found that perceived spatial separation between target speech and interfering speech only improved performance when TMRs were negative. For tasks that involved highly positive TMRs, no benefit of perceived spatial separation was found.

Interestingly, it appears that pitch does play some role in masking release: when the F_0 of the masker was shifted away from that of the target, masking release increased in both the LP and HP conditions. However, the improvement due to the pitch difference was the same in both conditions. This suggests that although the pitch of unresolved high-order harmonics is weaker than that for low-order harmonics, it is still sufficient for source segregation in dynamic situations, such as speech, where the moment-by-moment F_0 differences between the sources are typically much larger than the F_0 DL.

The conclusion that the pitch from unresolved harmonics may be sufficient for speech segregation, at least in certain circumstances, is consistent with basic psychoacoustic studies of stream segregation using only unresolved harmonics (Vliegen *et al.*, 1999; Vliegen and Oxenham, 1999). However, it seems at odds with the conclusions of Deeks and Carlyon (2004) and Qin and Oxenham (2005). Both these studies used different forms of envelope vocoders to study the role of pitch perception in speech segregation. Deeks and Carlyon (2004) used six-band envelope vocoders, excited by pulse trains with different F_0 s (or rates). They found no benefit of presenting the masker and target speech on different F_0 s in separate channels. However, their stimuli were much more impoverished than those used here: only three broad frequency bands were available for the target. Although this may be a realistic scenario for cochlear-implant users, it is not possible to conclude that it was simply the lack of resolved harmonics that led to poor performance and no benefit of F_0 differences. Qin and Oxenham (2005) found that the poorer F_0 discrimination produced by noise-excited envelope vocoding resulted in an inability to make use of F_0 differences between two simultaneously presented synthetic vowel stimuli. There are at least two reasons why they did not observe a benefit of F_0 difference while we did. First, modulated-noise stimuli from an envelope vocoder produce an even poorer pitch percept than unprocessed unresolved harmonics, meaning that the pitch information in the Qin and Oxenham (2005) study was more impoverished than that in the present study. Second, it may be that pitch information is more beneficial in sequential situations than in simultaneous

ones. For synchronously gated pairs of vowels, F_0 information from two sources must be extracted simultaneously; for more natural speech situations, much can be gained from “glimpsing” in time, as well as frequency, such that pitch differences across time can be used to form perceptual streams corresponding to the two sources.

The studies of masking release by Scott *et al.* (2001) and Elangovan and Stuart (2005) came to somewhat different conclusions from the ones reached here. Comparing masking by a steady noise with masking by a randomly interrupted noise masker, Scott *et al.* (2001) found that LP filtering reduced masking release, whereas Elangovan and Stuart (2005) found that HP filtering did not. Their results were discussed in terms of potentially poorer temporal resolution at low frequencies. There are several differences between our study and theirs. First, single-word recognition was tested, as opposed to sentences in the present study. Second, their noise was gated on and off abruptly, whereas we used either a single talker or a noise modulated by the broadband envelope of a single talker. It may be that the limits of temporal resolution (if any) are more apparent in abruptly gated maskers because the limits of temporal resolution (as opposed to the temporal slopes of the stimuli) are more likely to govern performance for abruptly gated maskers than for maskers with more shallow envelope slopes. Finally, it appears from Fig. 2 in Elangovan and Stuart (2005) that the spectra of their maskers and targets were not as carefully matched as in the present study, leaving open the possibility that some of the differential effects of filtering were due to spectrally local differences in long-term TMR. Thus, although the conclusions differ from ours, their data do not necessarily contradict our findings.

The results of Freyman *et al.* (2001) are somewhat related to, and are also generally consistent with, our findings. They found that the release from masking obtained by introducing a perceived spatial separation between a two-talker masker and a target speaker was no greater (in fact, somewhat less) when the stimuli were HP filtered at 1500 Hz than when they were unfiltered. This suggests that the masker and target were no more confusable after HP filtering, despite the reduction in the availability of low-frequency information. However, the use of female speech (with a higher F_0) and the absence of LP masking noise make it difficult to rule out the possibility that some low-frequency resolved harmonics were available in the stimuli of Freyman *et al.* (2001).

Whatever the reasons underlying the loss of masking release in the present experiments, it seems clear that simply bandlimiting the stimuli can be enough to severely reduce the masking release experienced by normal-hearing subjects. This implies that the proportion of the speech spectrum that is audible may itself be a significant predictor of how much masking release a given hearing-impaired listeners will demonstrate (e.g., Bacon *et al.*, 1998), as well as the many other potential factors, such as frequency and temporal resolution, that have been investigated in the past (e.g., George *et al.*, 2006, 2007).

VI. CONCLUSIONS

In unfiltered conditions, substantial release from masking was found when a steady-noise masker was replaced by a single-talker interferer or a speech-modulated-noise masker with the same long-term spectrum. The amount of masking release was greatly reduced in both LP- and HP-filtered conditions. In contrast to predictions based on the hypothesized role of pitch and temporal fine structure in masking, masking release in LP conditions was not significantly greater than in the HP conditions in any of the experimental conditions. This suggests that the temporal fine structure of resolved harmonics does not play a crucial or unique role in masking release. It may instead be that the amount of perceptual redundancy in the target determines masking release, or that any stimulus manipulation (or hearing condition) that increases the TMR necessary for intelligibility in steady noise to levels higher than 0 dB results in a loss of masking release.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grant No. R01 DC 05216). We thank Christophe Micheyl, Neal Viemeister, and Koen Rhebergen for many useful discussions, and Christian Lorenzi, Richard Freyman, Christophe Micheyl, and two anonymous reviewers for helpful comments on an earlier version of this paper.

APPENDIX: F_0 DIFFERENCE LIMENS AND HARMONIC RESOLVABILITY WITH SPEECH-SHAPED HARMONIC TONE COMPLEXES

1. Methods

a. Stimuli

In this experiment F_0 DLs were measured for harmonic tone complexes that were spectrally shaped to match the long-term power spectrum of the speech materials used in the main experiments. Equal-amplitude harmonic tone complexes, extending up to 10 kHz, were passed through a filter to provide them with the same spectral envelope as sentences from the HINT (Nilsson *et al.*, 1994). Two values of F_0 were tested, 110 and 135 Hz, which represent the mean F_0 and one standard deviation above the mean F_0 of the target speech, respectively. After spectral shaping, the complexes were either presented without any additional filtering (BB condition), or were LP or HP filtered, using a fourth-order Butterworth filter. When used, the LP filter cutoff was 1200 Hz and the HP filter cutoff was 1500 Hz, as in the main speech experiments. The components of the complex tones were added either in sine phase or in random phase. In the case of random phase, the phase of each component was chosen independently and randomly with uniform distribution from 0 to 2π for each stimulus presentation.

In the LP and HP conditions, a complementary noise was added to reduce the possibility of off-frequency listening and the detection of distortion products. This noise had the same speech-shaped spectral envelope as the tone complex but was 12 dB lower than the nominal level of the complex (before LP or HP filtering). In the LP condition, the noise

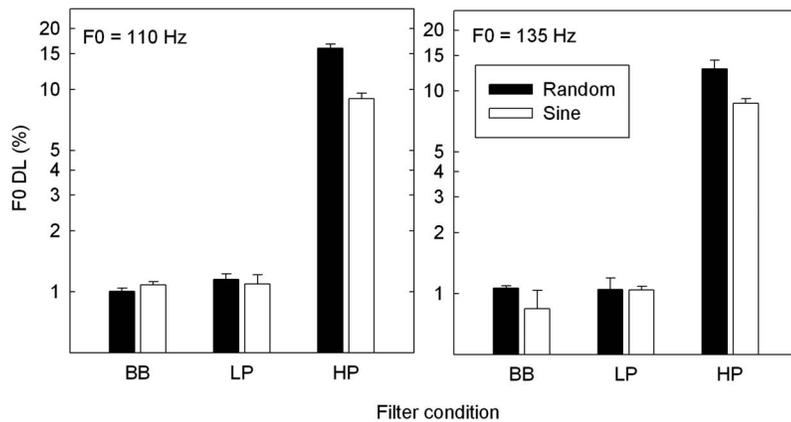


FIG. 5. F_0 DLs for tone complexes spectrally shaped to have the same spectral envelope as the target HINT sentences. Filter conditions include unfiltered (BB), LP filtered, and HP filtered. Solid bars represent conditions with components in random starting phase, and open bars represent conditions in sine starting phase. Error bars represent one standard error of the mean across subjects.

was HP filtered at 1200 Hz; in the HP condition, the noise was LP filtered at 1500 (fourth-order Butterworth filter in all cases). Thus the long-term spectral characteristics of the stimuli in this experiment were the same as the target and complementary noise in the main speech experiments. The tone complexes were 400 ms in duration, including 50-ms raised-cosine (Hanning) onset and offset ramps. The complementary noise was gated synchronously with the HP- or LP-filtered tone complex.

b. Procedure

Thresholds were measured using a two-interval two-alternative forced-choice procedure. In each trial, listeners were presented with two presentations of the tone complex, separated by a silent interval of 300 ms. The baseline F_0 ($F_{0_{\text{base}}}$) was roved over two semitones with uniform distribution across each trial (but not across intervals). The F_0 s in the two intervals were $F_{0_{\text{base}}}(1 + \Delta/100)^{1/2}$ and $F_{0_{\text{base}}}(1 + \Delta/100)^{-1/2}$, where Δ is the difference in F_0 between the two intervals, in percent. The listeners' task in each trial was to select the interval with the higher F_0 . Thresholds were estimated using a 3-down 1-up adaptive tracking procedure that estimates the 79.4% correct point of the psychometric function (Levitt, 1971). The initial value of Δ was 15.85%. This value was decreased after three consecutive correct responses and increased after each incorrect response. The transition from decreasing to increasing (or from increasing to decreasing) values defines a reversal. The initial step size was an increase or a decrease in Δ by a factor of 2.51. After two reversals, the step size decreased to a factor of 1.58, and after two more reversals, the step size was reduced to its final value of a factor of 1.26. A run terminated after six reversals at the final step size, and threshold was defined as the geometric mean value of Δ at those final six reversals.

The average overall level of the complex tone was set to 75 dB SPL but was roved over a uniform 6 dB range across intervals and trials. The level and F_0 roving were introduced to avoid the possibility that listeners learned to respond correctly based on cues other than F_0 (such as a small increase in loudness as a harmonic moved to the peak of the spectral envelope) and to simulate some aspects of the dynamic nature of speech, where the mean level may not be wholly representative. As in the speech experiments, the level of the complementary background noise was not roved.

The stimuli were presented diotically to listeners seated in a double-walled sound-attenuating booth. Stimuli were generated digitally and converted to voltage using a Lynx-Studio Lynx22 sound card with 24 bit resolution at a sampling rate of 22.05 kHz. The stimuli were either fed directly to HD580 headphones (Sennheiser) or were first passed through a headphone amplifier (TDT HB7). In both cases, levels were verified using a voltmeter and a sound pressure level meter coupled to the headphones using an artificial ear (Brüel and Kjær).

c. Subjects

Four normal-hearing listeners (all female; ages between 20 and 41) participated as subjects in this experiment. One was the second author; the others were paid for their services and had also participated in one of the speech experiments.

2. Results

The mean results, geometrically averaged across listeners, are shown in Fig. 5. In line with earlier studies using equal-amplitude harmonic tone complexes, F_0 DLs in the BB condition were all low ($\sim 1\%$) and showed no systematic effect of component phase relationship. In the LP conditions, F_0 DLs were similarly low, and also showed no phase dependencies. In contrast, F_0 DLs in the HP conditions were higher by nearly an order of magnitude, and showed a systematic phase effect, being higher (worse) in the random-phase conditions than in the sine-phase conditions. These results confirm that pitch coding accuracy for harmonic tone complexes with the same long-term spectral shape as the speech stimuli used in the main experiments is considerably worse in the HP-filtered conditions than in the LP-filtered or BB conditions. This is true for the F_0 corresponding to the mean F_0 of the sentences and for an F_0 one standard deviation above the mean. The poor thresholds and phase dependence in the HP-filtered conditions suggest that only unresolved harmonics were accessible to the subjects in the main speech experiments.

¹Some studies define the SRT as the TMR at which 50% of whole sentences (as opposed to words) are reported with no errors (e.g., Bacon *et al.*, 1998; Peters *et al.*, 1998).

²Results from the ANOVAs are quoted using the Greenhouse–Geisser correction for sphericity where appropriate (with original degrees of freedom cited).

³One subject had a threshold of 25 HL at 8000 Hz in one ear and a threshold of 25 HL at 4000 Hz in the other ear. However, the subject's data did not deviate significantly from those of others in that group.

- Assmann, P. F. (1999). "Fundamental frequency and the intelligibility of competing voices," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, pp. 179–182.
- Assmann, P. F., and Summerfield, A. Q. (2004). "The perception of speech under adverse conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. Fay (Springer, New York), pp. 231–308.
- Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," *J. Speech Lang. Hear. Res.* **41**, 549–563.
- Bernstein, J. G., and Oxenham, A. J. (2003). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?," *J. Acoust. Soc. Am.* **113**, 3323–3334.
- Bernstein, J. G., and Oxenham, A. J. (2006). "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," *J. Acoust. Soc. Am.* **120**, 3929–3945.
- 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.
- Boersma, P., and Weenink, D. (2008). "Praat: Doing phonetics by computer," <http://www.praat.org/> (Last viewed April 2008).
- Brox, J. P., and Nootboom, S. G. (1982). "Intonation and the perceptual separation of simultaneous voices," *J. Phonetics* **10**, 23–36.
- Brungart, D. S. (2001). "Informational and energetic masking effects in the perception of two simultaneous talkers," *J. Acoust. Soc. Am.* **109**, 1101–1109.
- Carlyon, R. P. (1996). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," *J. Acoust. Soc. Am.* **99**, 517–524.
- Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," *J. Acoust. Soc. Am.* **114**, 2913–2922.
- Deeks, J. M., and Carlyon, R. P. (2004). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," *J. Acoust. Soc. Am.* **115**, 1736–1746.
- Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source on the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.* **74**, 739–743.
- Elangovan, S., and Stuart, A. (2005). "Interactive effects of high-pass filtering and masking noise on word recognition," *Ann. Otol. Rhinol. Laryngol.* **114**, 867–878.
- 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.
- French, N. R., and Steinberg, J. C. (1947). "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Am.* **19**, 90–119.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). "Spatial release from informational masking in speech recognition," *J. Acoust. Soc. Am.* **109**, 2112–2122.
- Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2008). "Spatial release from masking with noise-vocoded speech," *J. Acoust. Soc. Am.* **124**, 1627–1637.
- George, E. L., Festen, J. M., and Houtgast, T. (2006). "Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **120**, 2295–2311.
- George, E. L., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., and Houtgast, T. (2007). "Auditory and nonauditory factors affecting speech reception in noise by older listeners," *J. Acoust. Soc. Am.* **121**, 2362–2375.
- Gnansia, D., Jourdes, V., and Lorenzi, C. (2008). "Effect of masker modulation depth on speech masking release," *Hear. Res.* **239**, 60–68.
- 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.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **17**, 225–246.
- Jin, S. H., and Nelson, P. B. (2006). "Speech perception in gated noise: The effects of temporal resolution," *J. Acoust. Soc. Am.* **119**, 3097–3108.
- Kwon, B. J., and Turner, C. W. (2001). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," *J. Acoust. Soc. Am.* **110**, 1130–1140.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869.
- Micheyl, C., Bernstein, J. G., and Oxenham, A. J. (2006). "Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise," *J. Acoust. Soc. Am.* **120**, 1493–1505.
- 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.
- Miller, G. A., and Liclclider, J. C. R. (1950). "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* **22**, 167–173.
- Nelson, P. B., and Jin, S. H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Nilsson, M., Soli, S., 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.
- Northern, J., and Downs, M. (2002). *Hearing in Children* (Lippincott, Williams & Wilkins, New York).
- Peters, R. W., Moore, B. C. J., and Baer, T. (1998). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," *J. Acoust. Soc. Am.* **103**, 577–587.
- 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.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2005). "Release from informational masking by time reversal of native and non-native interfering speech," *J. Acoust. Soc. Am.* **118**, 1274–1277.
- Rhebergen, K. S., Versfeld, N. J., and Dreschler, W. A. (2006). "Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise," *J. Acoust. Soc. Am.* **120**, 3988–3997.
- Scott, T., Green, W. B., and Stuart, A. (2001). "Interactive effects of low-pass filtering and masking noise on word recognition," *J. Am. Acad. Audiol.* **12**, 437–444.
- 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.
- Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (2007). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.* **122**, 1069–1078.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Summers, V., and Leek, M. R. (1998). "F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss," *J. Speech Lang. Hear. Res.* **41**, 1294–1306.
- Summers, V., and Molis, M. R. (2004). "Speech recognition in fluctuating and continuous maskers: Effects of hearing loss and presentation level," *J. Speech Lang. Hear. Res.* **47**, 245–256.

- Takahashi, G. A., and Bacon, S. P. (1992). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," *J. Speech Hear. Res.* **35**, 1410–1421.
- Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (1999). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," *J. Acoust. Soc. Am.* **106**, 938–945.
- Vliegen, J., and Oxenham, A. J. (1999). "Sequential stream segregation in the absence of spectral cues," *J. Acoust. Soc. Am.* **105**, 339–346.
- Whitmal, N. A., Poissant, S. F., Freyman, R. L., and Helfer, K. S. (2007). "Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience," *J. Acoust. Soc. Am.* **122**, 2376–2388.