Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept

Andrew King, Léo Varnet, Christian Lorenzi

Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept

Andrew King,[a] Léo Varnet, and Christian Lorenzi
Laboratoire des systèmes perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, École normale supérieure, Université Paris Sciences & Lettres, 29 rue d'Ulm, 75005 Paris, France

(Received 31 July 2018; revised 25 January 2019; accepted 25 February 2019; published online 24 April 2019)

Frequency modulation (FM) is assumed to be detected through amplitude modulation (AM) created by cochlear filtering for modulation rates above 10 Hz and carrier frequencies (f_c) above 4 kHz. If this is the case, a model of modulation perception based on the concept of AM filters should predict masking effects between AM and FM. To test this, masking effects of sinusoidal AM on sinusoidal FM detection thresholds were assessed on normal-hearing listeners as a function of FM rate, f_c, duration, AM rate, AM depth, and phase difference between FM and AM. The data were compared to predictions of a computational model implementing an AM filter-bank. Consistent with model predictions, AM masked FM with some AM-masking-AM features (broad tuning and effect of AM-masker depth). Similar masking was predicted and observed at $f_c = 0.5$ and 5 kHz for a 2 Hz AM masker, inconsistent with the notion that additional (e.g., temporal fine-structure) cues drive slow-rate FM detection at low f_c. However, masking was lower than predicted and, unlike model predictions, did not show beating or phase effects. Broadly, the modulation filter-bank concept successfully explained some AM-masking-FM effects, but could not give a complete account of both AM and FM detection.

https://doi.org/10.1121/1.5094344

[VMR] Pages: 2277–2293

I. INTRODUCTION

Many natural sounds such as communication calls or speech convey strong modulations in amplitude and frequency (e.g., Steeneken and Houtgast, 1980; Hsu *et al.*, 2004; Sheft *et al.*, 2012; Varnet *et al.*, 2017). As for speech, frequency-modulation (FM) cues correspond to relatively slow (<5 Hz) fluctuations in the fundamental frequency that have been shown to play a critical role in the ability to identify speech when presented against concurrent speech sounds (Binns and Culling, 2007). Consistent with this idea, there is evidence of correlations between individual auditory sensitivity to FM at 2 or 5 Hz and speech-recognition capacities for normal-hearing and hearing-impaired listeners when target speech is masked by concurrent speech sounds (Ruggles *et al.*, 2011; Johannesen *et al.*, 2015).

Over the last few decades, a wealth of psychophysical studies were conducted to explore the low-level, auditory mechanisms responsible for FM sensitivity. The general consensus is that for low carrier frequencies (<1–4 kHz) and at slow modulation rates (<5–10 Hz), FM detection relies mainly on the use of temporal fine-structure (TFS) cues (for a review, see Moore, 2014). These cues correspond to the fast oscillations in instantaneous frequency evoked by FM tones at the output of cochlear filters that are encoded via the phase-locked activity of auditory neurons in the auditory periphery or lower brainstem (Paraouty *et al.*, 2018). For higher carrier frequencies and at faster modulation rates, FM detection is mainly based on the use of excitation-pattern cues. These cues correspond to temporal-envelope cues resulting from the differential attenuation produced by cochlear filters (Zwicker, 1956; Saberi and Hafter, 1995). These fluctuations are encoded by the changes in the mean firing rate of auditory neurons, or in other words, by a rate-place (tonotopic) code (Moore and Sek, 1995; Sek and Moore, 1995; Moore and Sek, 1996; Paraouty *et al.*, 2018; for a recent review, see Ernst and Moore, 2010, 2012).

This two-mechanism theory based on TFS and temporal-envelope cues explains, for instance, the specific pattern of interference produced by an amplitude modulation (AM) superimposed onto all stimuli when measuring frequency-modulation detection thresholds (FMDTs) for the same modulation rates (Moore and Sek, 1996; Paraouty *et al.*, 2016; Paraouty and Lorenzi, 2017; Paraouty *et al.*, 2018). The AM is intended to disrupt temporal-envelope cues for FM detection by introducing fluctuations in excitation level that are uninformative about the FM. Several studies (e.g., Moore and Sek, 1996; Ernst and Moore, 2010) repeatedly showed that the added AM adversely affects FMDTs and, for carrier frequencies below 4 kHz, the adverse effect increases with increasing modulation rate, consistent with the idea that temporal-envelope cues play a greater role for higher modulation rates. For a carrier frequency (f_c) of 6 kHz, the adverse effect of the added AM is similar for all modulation rates, consistent with the idea that, for very high carrier frequencies, temporal-envelope cues dominate for all modulation rates.

[a]Electronic mail: andrewk@dtu.dk

This two-mechanism theory was recently challenged by a correlational study investigating individual differences in FM and AM sensitivity (Whiteford and Oxenham, 2015). The outcome of this study based on a large cohort of 100 normal-hearing participants showed that FMDTs were strongly correlated with amplitude modulation detection thresholds (AMDTs). However, correlations were not stronger between psychophysical measures assumed to reflect the use of TFS information (i.e., between slow-rate FM sensitivity and sensitivity to interaural time differences), or between measures assumed to reflect the use of temporal-envelope information (i.e., fast-rate FM sensitivity and forward-masking patterns). A follow-up study with listeners of a wider age range also concluded that FM detection cannot be unambiguously ascribed to TFS coding (Whiteford *et al.*, 2017). These findings indicate that further work is warranted to challenge the two-mechanism theory proposed for FM detection.

The goal of the present study is to test the validity of this theory by exploring *systematically* masking effects between AM and FM using the framework of the modulation filter-bank concept (Dau *et al.*, 1997). Within this concept is the idea that the auditory system is not only tuned in the audio frequency domain, but also in the AM domain. Assuming this to be the case, any masking created by AM on FM detection should be explicable by the filtering effects caused by a modulation filter-bank, thereby providing a unified account of AM and FM auditory processing. For high carrier frequencies (>1–4 kHz) and at all modulation rates, and for low carrier frequencies and at rates faster than about 5–10 Hz, FM is assumed to be encoded as AM (i.e., temporal-envelope cues). As indicated above, this idea is supported by the repeated observation of an increase in FMDTs when FM is presented with a sinusoidal AM at the same rate or with an AM noise centered at the FM rate under the stimulus configurations noted above (e.g., Moore and Sek, 1996; Ernst and Moore, 2010; Paraouty *et al.*, 2016). If this idea is correct, the masking effects between AM and FM should follow closely the pattern described previously for "AM-masking-AM" conditions (Houtgast, 1989; Bacon and Grantham, 1989; Strickland and Viemeister, 1996; Ewert and Dau, 2000) and show the following four important features:

(1) Frequency selectivity or "tuning" (Houtgast, 1989; Bacon and Grantham, 1989; Strickland and Viemeister, 1996; Ewert and Dau, 2000; Lorenzi *et al.*, 2001): The greatest amount of AM (or modulation) masking typically occurs when the target AM rate is near the masker AM rate. Importantly, modulation masking extends over a fairly broad range, and the −3 dB bandwidth of the masking patterns is about 1 octave.
(2) Effects of masking AM depth (Bacon and Grantham, 1989; Strickland and Viemeister, 1996): The greater the modulation depth of the masker, the greater the modulation masking. More precisely, modulation masking grows linearly with the modulation index of the masker (in dB *versus* dB).
(3) Phase effects (Bacon and Grantham, 1989; Strickland and Viemeister, 1996; Lorenzi *et al.*, 1999): AMDTs for a target AM vary as a function of the masking AM phase when the target AM rate is half or twice the masker AM rate. When the target AM rate is half the masker AM rate, the functions relating AMDTs to the target–masker phase difference have two peaks. These results were found with noise carriers, and it is still unclear whether pure-tone carriers would produce similar phase effects on AMDTs or not.
(4) Beating effects (Strickland and Viemeister, 1996; Lorenzi *et al.*, 2001; Millman *et al.*, 2002): AMDTs decrease (improve) abruptly when the masker AM rate is only 2–4 Hz greater than target AM rate. This reduction in modulation masking suggests that listeners use a low-rate temporal-envelope beat cue (a cyclic increase and decrease in the modulation depth due to the modulations cyclically moving in and out of phase) when the difference in rate between the target and masker modulations is small. The beat cue has a rate equal to this difference.

To test this, FMDTs were measured in several conditions for a group of normal-hearing adult listeners. The first experiment was designed to examine the tuning of masking and beating effects in the modulation domain. Here, FMDTs for a sinusoidal FM were measured at modulation rates between 2 and 64 Hz for both a low (0.5 kHz) and a high (5 kHz) pure-tone carrier. The carrier was either unmodulated in amplitude or modulated sinusoidally in amplitude at 2 or 16 Hz at a modulation depth of 50%. To measure the use of beats between an AM masker of 16 Hz with FM target of 14 and 18 Hz, two stimulus durations were used (0.5 and 1 s), which provided 1 or 2 cycles of the 2-Hz beat, respectively. We reasoned that two beat cycles should provide a better cue than one, thus producing more release from masking (as observed for AM-masking-AM by Millman *et al.*, 2002). The difference in masking between these two duration conditions should reveal any beating effect. In addition, AMDTs for a sinusoidal AM were measured at modulation rates of 2, 4, 8, 16, and 32 Hz for both a low (0.5 kHz) and a high (5 kHz) pure-tone carrier. Stimulus duration was set to either 0.5 or 1 s.

The second experiment examined the effect varying the masker modulation depth. This experiment repeated the first, except with a masker modulation depth set at 25% and stimulus duration set at 1 s. The third experiment examined the effect of varying systematically the phase relationship between the FM target and the AM masker in 45° steps. Here, FMDTs for a sinusoidal FM were measured at two modulation rates: 16 Hz with a 5 kHz pure-tone carrier modulated sinusoidally in amplitude at 8, 16, or 32 Hz; and 2 Hz with a 0.5 kHz pure-tone carrier sinusoidally modulated in amplitude at 2 or 4 Hz. The masker modulation depth was fixed to 50%, and stimulus duration was set to 1 s.

It is important to note that if FM is encoded as AM at high modulation rates and carrier frequencies, then the AM induced by the FM would have a different phase to the FM, depending on where (i.e., in which cochlear filter) the conversion from FM to AM occurs. Exact parameters such as this—and others such as the depth of the AM induced by the FM—are unknown.

If FM is actually encoded as AM, it should be possible to give a unified account of AM-masking-AM and AM-masking-FM within the framework of the modulation filter-bank concept (Dau *et al.*, 1997). A computational model based on the

modulation-filterbank concept and a template-matching decision strategy (Dau *et al.*, 1997; Wallaert *et al.*, 2017) was developed to account for the FM data. The model incorporated stages simulating temporal-envelope processing by low-level sensory mechanisms (cochlear filtering, instantaneous amplitude compression, adaptation), mid-level processes (bandpass AM filtering), and higher-level non-sensory processes (internal noises, memory decay in the temporal-envelope domain, and template matching). The model was used to predict the effects of modulation rate, depth, stimulus duration, and modulation phase on FMDTs in the presence of an AM masker on the sole basis of temporal-envelope cues resulting from FM-to-AM conversion at the outputs of the cochlear filters. Therefore, the model was expected to produce positive and quantitative predictions for the four features of modulation masking detailed above.

In summary, in the conditions for which FM is assumed to be encoded using a temporal-envelope code, the general features of AM-masking-AM listed above were expected (masking, tuning, beating, depth, and phase effects). However, for the conditions in which FM is assumed to be encoded using a temporal fine structure code (slow-rate FM and low f_c), little masking between AM and FM, if any, was expected.

II. EXPERIMENT 1

A. Methods

1. Listeners

Ten listeners (seven female) between 20 and 30 years old (mean 25.3 years) participated. Audiometric thresholds were tested at 0.25, 0.5, 1, 2, 4, 6, and 8 kHz using a Madsen Itera II (Otometrics, Taastrup, DK) audiometer, following the recommended procedure of the British Society of Audiology (British Society of Audiology, 2011). All listeners had audiometric thresholds at or below 20 dB hearing level (HL), for both ears, at all frequencies tested. All listeners also had pure-tone absolute thresholds of less than 20 dB sound pressure level (SPL) at 0.5 and 5 kHz (see below). Listeners were recruited through a national research participation database (the *Relais d'information sur les sciences de la cognition*) and from the students and staff of the laboratory, including the first author (S01). All participants gave informed consent. The study was approved by the local ethical committee of University Paris Descartes (IRB 20143200001072).

A priori power analysis, based on an effect size similar to the release from modulation masking due to beating between the target and masker found by Millman *et al.* (2002) suggested a sample size of ten listeners should achieve a statistical power of 0.8.

2. Stimuli

All stimuli were generated digitally at a sample rate of 48 kHz using MATLAB R2013b (The Mathworks, Natick, MA) and, using the built-in "audioplayer" command, sent to an audioengine D3 digital-to-analog audio converter (Austin, TX) for conversion at a bit depth of 2^{24}. The analog signals were presented to listeners with Sennheiser HD600 headphones (Old Lyme, CT) within a double-walled sound-proof booth via a wall patch. HD600 headphones have a diffuse-field equalized frequency response. Hence the response is not flat across frequency, but varies smoothly. These variations in level across frequency could transform FM into AM before the stimulus even reaches the ear. Fluctuations in level with frequency were measured with a sound level meter (Brüel and Kjær type 2250) and an artificial ear (Brüel and Kjær type 4153) with pure tones at 0.5 and 5 kHz modulated in frequency at 0.1 Hz. The slow FM rate allowed the temporal integration window of the sound level meter to update (at a rate of 1 Hz) as the FM tone changed in frequency. Deviations in level were tested at frequency excursions well above threshold and approximately around threshold ($\pm 10\%$ and $\pm 0.5\%$ of the f_c, respectively). For frequency excursions of $\pm 10\%$ of the f_c, level deviations were 0.9 and 1.2 dB for $f_c = 0.5$ and 5 kHz, respectively. For frequency excursions of $\pm 0.5\%$ of the f_c, level deviations were 0.1 dB for both carrier frequencies.

Absolute thresholds for the detection of 0.5 and 5 kHz pure tones (the carrier frequencies used for the modulation detection tasks) were tested by a 3-interval, 3-alternative forced-choice (AFC) detection task. The 3-AFC task consisted of two intervals of silence and one containing a tone at an adaptively tracked level, all in a random order. The intervals were 300 ms separated by 300 ms pauses. The tones had 20 ms raised-cosine ramps.

Sinusoidal AM and FM detection were tested with two signal durations, 0.5 and 1 s, in a 2-interval, 2-AFC task. The target and reference intervals were randomly ordered. Stimulus intervals were onset and offset with 20 ms raised-cosine ramps. The inter-stimulus interval was 400 ms and there was a 350 ms interval between the participant response and the beginning of the next trial. The signals were equalized by their root-mean-square (rms) amplitude, and played out of the right headphone at level that randomly varied (uniform distribution) around 60 dB SPL by $\pm$ 3 dB between intervals and across trials.

AM tones were generated by the formula given in Eq. (1),

$$AM\ tone = \left[1 + 10^{m/20} \sin\left(2\pi f_{AM} t + \Phi_{AM}\right)\right] \cdot \sin\left(2\pi f_c t + \varphi\right), \quad (1)$$

where m is the AM depth in dB (re 100% modulation), f_{AM} is the AM rate in Hz, t is the time-sample vector, Φ_{AM} is the modulation phase, and φ is the carrier phase. For AM detection, m was adaptively tracked to obtain the threshold (referenced to a pure tone at the same f_c, see Sec. II A 3), whilst f_{AM} was 2, 4, 8, 16, or 32 Hz, Φ_{AM} varied randomly $[U(0,2\pi)]$ on each trial, f_c was 0.5 or 5 kHz, and φ was varied randomly $[U(0,2\pi)]$ on every stimulus (i.e., different for target and reference).

FM tones were generated by the formula given in Eq. (2),

$$FM\ tone = \sin[2\pi f_c t + \varphi + (\Delta f / f_{FM}) \cdot \sin(2\pi f_{FM} t + \Phi_{FM})], \quad (2)$$

where Δf is the frequency excursion from f_c (in Hz) and f_{FM} is the FM rate in Hz, and Φ_{FM} is the modulation phase. For

FM detection without AM, Δf was adaptively tracked (also referenced to a pure tone at the same f_c). f_{FM} was 2, 4, 8, 14, 16, 18, 32, or 64 Hz, f_c was 0.5 or 5 kHz, and Φ_{FM} was varied randomly [$U(0,2\pi)$] on each trial and φ was varied randomly [$U(0,2\pi)$] on every stimulus.

For FM detection with AM masking, the stimuli were created by the formula in Eq. (3),

$$AM\text{–}FM\ tone = \left[1 + 10^{m/20}\sin\left(2\pi f_{AM}t + \Phi_{AM}\right)\right] \cdot FM\ tone, \tag{3}$$

where *FM tone* is derived from Eq. (2). The target combined AM–FM tone had a f_{AM} of either 2 or 16 Hz, an m of −6.02 dB (50%), an f_{FM} of 2, 4, 8, 14, 16, 18, 32, or 64 Hz, and an f_c of 0.5 or 5 kHz. In the target stimulus, the Φ_{AM} and Φ_{FM} were equal {which randomly varied [$U(0,2\pi)$] on each trial}, whilst φ varied randomly [$U(0,2\pi)$] on every stimulus (both target and reference). Again, Δf was adaptively tracked, but this time referenced to an AM tone [using Eq. (1)] at the same f_c, f_{AM}, m, and Φ_{AM} as the target stimulus.

3. Procedure

The 3-AFC absolute detection and 2-AFC modulation detection tasks were conducted in the MATLAB environment using the psychoacoustics framework AFC (version 1.40.1; Ewert, 2013). Thresholds were estimated with adaptive staircases that attempted to converge on 71% correct detection using transformed 2-down, 1-up method (e.g., Levitt, 1971). Two consecutive correct responses resulted in a reduction in the dependent variable (DV) for the subsequent trial, whereas any incorrect response resulted in an increase in the DV for the subsequent trial (up to maximum limits). Listeners were given visual feedback after each trial. If the maximum limit was reached three times in any given staircase, that staircase was terminated and skipped. The maximum limit was 0 dB for AM detection and 10% of the f_c for FM detection. This happened once each for P02 and P04 in the training block for FM detection with AM masking, once each for P04 and P07 in the AM detection training block, and once for S08 during the AM detection main, test block. Although skipped tracks were not repeated, multiple staircases (no fewer than three, including training) were completed for each condition.

For the absolute detection task, listeners were instructed to select the interval containing the tone. The DV (stimulus level) started at 45 and 50 dB SPL for 0.5 and 5 kHz, respectively. Step sizes for increments and decrements of level were 8 dB until the first reversal from increments to decrements, and then 2 dB for six reversals thereafter. The mean level of these last six reversals was taken as the threshold.

For the AM detection task, listeners were instructed to select the interval containing the modulation or fluctuation in amplitude. The DV (m) started at −6 dB for all conditions. Step sizes for increments and decrements of m were 4 dB until the first reversal from increments to decrements, then 2 dB for two more reversals, and 1 dB for eight reversals thereafter. The mean of m at these last eight reversals was taken as the threshold.

For the FM detection tasks (with and without AM masking), listeners were instructed to select the interval containing the fluctuation in pitch. The DV (Δf) started at 4% of the f_c (20 and 200 Hz for 0.5 and 5 kHz carriers, respectively). Step sizes for increments and decrements of frequency excursion were multiplicative factors of 1.58 until the first reversal from increments to decrements, then 1.26 for two more reversals, and 1.12 for eight reversals thereafter. The geometric mean of Δf at these last eight reversals was taken as the threshold.

Listeners completed one staircase for each condition as training (totaling 116 staircases), beginning with AM detection, then FM detection without AM maskers, and finally FM detection with AM maskers. For each task, the conditions were randomly ordered. After this, listeners completed three staircases for each condition of each task. The AM detection task was completed first and the FM detection second, with the conditions with and without AM maskers mixed together in a random order. For each task, all conditions were presented before any condition was repeated.

The experiment took listeners approximately 25 h to complete, over multiple sessions. Sessions were generally between 1.5 and 2 h long and could be stopped at the end of any staircase. Each session began with two absolute threshold staircases, one at 0.5 kHz and one at 5 kHz, both in the right (test) ear, to ensure that the listener's hearing had not changed substantially between sessions (e.g., due to recent noise exposure). The non-training sessions also included a warm-up staircase of the current modulation detection task, after the two absolute threshold staircases.

B. Results

1. AM detection thresholds

As plotted in the upper two panels of Fig. 1, mean AMDTs were very similar for $f_c = 0.5$ kHz and $f_c = 5$ kHz. The thresholds were highest (worst) for the 2-Hz AM rate for both 1.0-s (−17 dB) and 0.5-s (−11 dB) duration. For 0.5-s stimuli, AMDTs decreased from −18 dB at 4 Hz to −22 dB at 8 Hz, then remained constant up to 32 Hz. For 1-s stimuli, AMDTs were relatively constant at −24 dB from 4 to 32 Hz. AMDTs for the 1-s duration were better than threshold for the 0.5-s duration and this was most pronounced at 2 and 4 Hz. For instance, at 2 Hz two modulation cycles, rather than one, detection improved by 6 dB. A repeated-measures analysis-of-variance (ANOVA) of the AMDTs with factors for AM rate, duration and f_c confirmed significant effects of both AM rate [$F(4,36) = 68.16$, $p < 0.001$, $\eta p^2 = 0.88$] and duration [$F(1,9) = 184.01$, $p < 0.001$, $\eta p^2 = 0.95$] and a significant interaction between them [$F(4,36) = 41.59$, $p < 0.001$, $\eta p^2 = 0.82$]. There was no significant effect of f_c [$F(1,9) = 1.20$, $p = 0.30$, $\eta p^2 = 0.12$] and none of the interactions with f_c were significant.

2. FM detection thresholds

In Fig. 1, the second row (from top) of panels shows the group-mean FMDTs for the conditions without AM applied to the carrier stimulus. Geometric mean FMDTs generally ranged from 0.005 to 0.012 peak-to-peak deviations as a

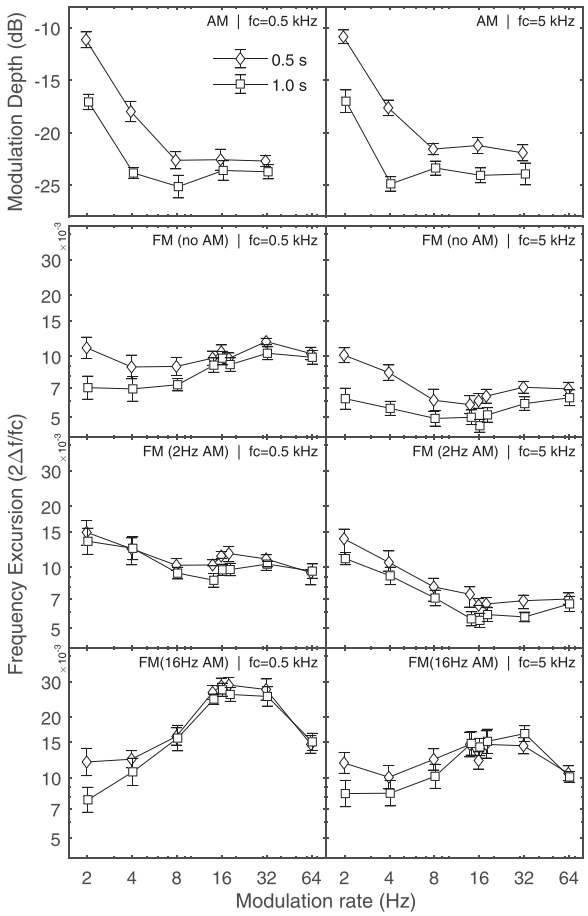

FIG. 1. AMDTs (top row), FMDTs without AM masking (second row from top), FMDTs with a 2-Hz AM masker (third row from top), and FMDTs with a 16-Hz AM masker (bottom row) plotted as a function of modulation rate, for stimuli of 0.5 s (diamonds) and 1 s (squares) durations and a carrier frequency (f_c) of 0.5 kHz (left) and 5 kHz (right), with ±1 standard error (SE) bars.

proportion of f_c. In contrast with the AM data, mean FMDTs were different for the two carrier frequencies. They were generally higher when $f_c = 0.5$ kHz than when $f_c = 5$ kHz. For $f_c = 0.5$ kHz and the 1-s duration, FMDTs slightly increased with FM rate. For $f_c = 0.5$ kHz, and the 0.5 s duration, FMDTs were roughly constant across FM rates. FMDTs also remained roughly constant across FM rates for $f_c = 5$ kHz and the 1 s duration. For $f_c = 5$ kHz and the 0.5 s duration, FMDTs decreased with increasing FM rate up to 8 Hz. Across all FM rates, and for both carriers, FMDTs were higher for the 0.5-s duration stimuli than for the 1-s stimuli. This difference was more pronounced at lower FM rates than at higher FM rates.

The third row of panels of Fig. 1 shows the group-mean FMDTs with a 2-Hz AM masker applied to the carrier. Again, FMDTs were generally higher when $f_c = 0.5$ kHz than when $f_c = 5$ kHz. Above an FM rate of 16 Hz, FMDTs are similar to those without an AM masker, but below 14 Hz, FMDTs increase with decreasing FM rate and are elevated compared to those without an AM masker. This change in thresholds with FM rate is relatively equal for both stimulus durations. However, due to the difference in thresholds between 0.5- and 1-s stimulus durations at slow FM rates without an AM masker, the masking observed at 2 and 4 Hz FM rates is greater for the 1 s than the 0.5 s duration. This can be seen more clearly in the top two panels of Fig. 2, where the ratios of thresholds with a 2-Hz AM masker (FM_{wAM}) to thresholds without AM masking (FM_{noAM}) are plotted in a similar format to Fig. 1. These data (in the top panels of Fig. 2) also show masking is strongest at 2- and 4-Hz FM rates (ratios around 1.9 for 1-s stimuli and 1.4 for 0.5-s stimuli) and disappears above 14 Hz. The horizontal gray dashed lines at a ratio of 1 mark the point of no masking, i.e., thresholds with and without AM are equal.

The bottom row of panels of Fig. 1 show the FMDTs with a 16-Hz AM masker applied to the carrier. Again, FMDTs were generally higher when $f_c = 0.5$ kHz than when $f_c = 5$ kHz. However, this is most pronounced at or above an FM rate of 8 Hz. FMDTs peaked at FM rates of 16 to 32 Hz at values of approximately 0.03 for $f_c = 0.5$ kHz and 0.015 for $f_c = 5$ kHz. Below an FM rate of 8 Hz, FMDTs were lower for 1 s stimuli than the 0.5 stimuli, but at 8 Hz and above, FMDTs were similar for both stimulus durations. The bottom two panels of Fig. 2 show the ratios of the thresholds with a 16-Hz AM masker to thresholds without AM masking. They show that masking peaked at an FM rate of 16 Hz, with a peak ratio of about 2.8 to 3. There was little effect of duration for $f_c = 0.5$ kHz and slightly more masking for 1-s duration stimuli than 0.5-s duration stimuli for $f_c = 5$ kHz.

An ANOVA with four factors was performed on the log-transformed thresholds ($2\Delta f$ divided by the f_c) using the afex R-package (Singmann *et al.*, 2017). The factors were [FM rate (8 levels: 2, 4, 8, 14, 16, 18, 32, and 64 Hz); f_c (2 levels: 0.5 and 5 kHz); stimulus duration (2 levels: 0.5 and 1 s); and AM masker rate (3 levels: 0, or no AM, 2, and 16 Hz)], with all interactions included in the model (type III sum of squares). The main effects of f_c, stimulus duration and AM masker rate were all significant [f_c: $F(1,9) = 38.44$, $p < 0.01$, $\eta p^2 = 0.81$; duration: $F(1,9) = 56.16$, $p < 0.001$, $\eta p^2 = 0.86$; AM masker rate: $F(1,9) = 164.87$, $p < 0.001$,

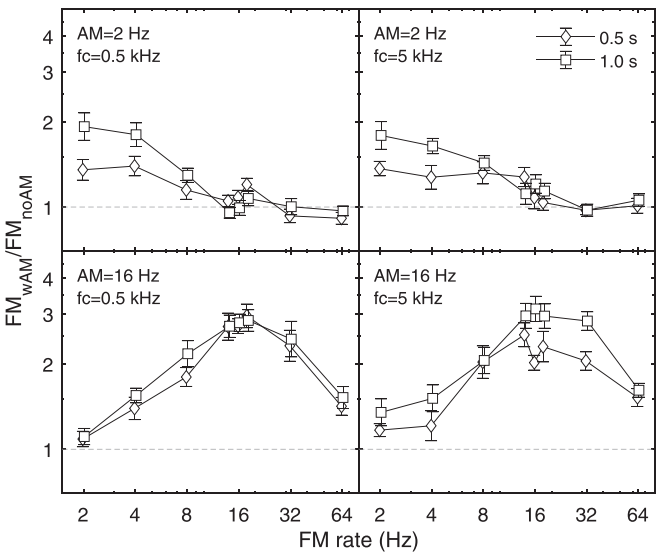

FIG. 2. Masking patterns by FM rate, for a 2-Hz AM masker (top) and a 16-Hz AM masker (bottom). Masking defined as FMDTs with an AM masker (modulation depth = 50%) divided by FMDTs without an AM masker. Symbols are as in Fig. 1 with ±1 SE bars.

$\eta p^2 = 0.95$]. FMDTs were generally higher with an f_c of 0.5 than 5 kHz, generally higher with a stimulus duration of 0.5 than 1 s, and generally lowest with no AM masker and highest with a 16-Hz AM masker. However, these main effects are marginal to several significant interactions. Namely, the two-way interactions between each of these factors and the effect of FM rate [f_c*FM rate: $F(7,63) = 10.02$, $p < 0.001$, $\eta p^2 = 0.53$; duration*FM rate: $F(7,63) = 8.50$, $p < 0.001$, $\eta p^2 = 0.49$; AM masker*FM rate: $F(14\,126) = 56.64$, $p < 0.001$, $\eta p^2 = 0.86$] and the interaction between AM masker rate and duration [$F(2,18) = 15.30$, $p < 0.01$,$\eta p^2 = 0.63$]. The latter three interactions were in turn marginal to the three-way interaction between AM masker, duration and FM rate [$F(14\,126) = 2.76$, $p < 0.05$, $\eta p^2 = 0.23$]; duration appears to have more of an effect without an AM masker for slower FM rates than faster ones, whereas with a 2-Hz AM and a 16-Hz AM there is little effect of duration, particularly when AM and FM rate are close (where masking is greatest). However, regardless of duration, the FM rate at which thresholds are greatest clearly depends on the rate of the AM masker. The main effect of FM rate just failed to be significant [$F(7,63) = 3.03$, $p = 0.057$, $\eta p^2 = 0.25$] and all other interactions were not significant.

The masking patterns, particularly for the 1-s stimuli, exhibited *tuning*; in other words, for both AM masker rates, masking culminated when FM and AM-masker rates are equal. However, this tuning is relatively *broad* at both AM masker rates; masking decreased by half only when AM and FM rates differed by one or two octaves. Masking at the masker rate was greater for the 16-Hz AM masker than for the 2-Hz AM masker. An ANOVA of the log-transformed masking ratios for the on-frequency conditions only ($f_{FM} = f_{AM} = 2$ or 16 Hz) found a significant difference between the two masker rates [$F(1,9) = 39.94$, $p < 0.001$], and a significant effect of duration [$F(1,9) = 26.22$, $p < 0.001$], but no significant effect of f_c or any significant interactions.

In summary, these analyses revealed significant masking between AM and FM. As expected, the masking was broadly tuned. Analyses also revealed greater masking for a 16-Hz AM-masker rate than a 2-Hz rate, and greater masking for the longer stimulus duration. Inconsistent with our initial expectations, these masking patterns were similar across the low and high carrier frequencies.

3. Beating effects

A release from masking was expected for the conditions with $f_{AM} = 16$-Hz, if the 2-Hz beat occurring between the 14- and 18-Hz FM and the 16-Hz AM masker facilitated detection of the FM target. In particular, it was expected to produce greater release from masking for the 1-s duration stimuli than the 0.5-s duration. Figure 2 shows that masking was not greatly reduced at FM rates of 14 and 18 Hz, compared to masking at 8, 16, or 32 Hz, for either stimulus duration or f_c. A three-way ANOVA was performed on the (log-transformed) masking ratios when $f_{AM} = 16$ Hz, with factors of f_c (0.5 and 5 kHz), duration (0.5 and 1 s) and FM rate (14, 16, and 18 Hz). It showed no significant effect of FM rate [$F(2,18) = 0.23$, $p = 0.80$, $\eta p^2 = 0.02$]. There was a significant effect of duration [$F(1,9) = 7.13$, $p < 0.05$, $\eta p^2 = 0.44$], but this appears to only occur for $f_c = 5$ kHz, where masking was generally less for the 0.5-s duration across FM rates than the 1-s duration. The significant interaction between f_c and duration corroborates this [$F(1,9) = 26.96$, $p < 0.001$, $\eta p^2 = 0.75$]. No other interactions were significant. In summary, in contrast to our initial expectations, the interaction between AM and FM did not produce beating effects that reduced masking when AM and FM are very close in rate, even for the conditions in which FM is assumed to be encoded using a temporal-envelope code.

III. EXPERIMENT 2: EFFECTS OF AM-MASKER DEPTH

In order to test whether the masking of FM by AM featured a dependency on the masker AM depth, a subset of listeners also performed FM detection across the same range of FM rates and the same f_c (0.5 and 5 kHz), but with the 2- and 16-Hz AM maskers at 25% depth, rather than 50%. The hypothesis was that more masking would be observed with a 50% AM masker depth than a 25% AM masker depth. Furthermore, this effect was expected to predominantly happen when the FM and AM rates were equal and less when the rates are very different.

A. Methods

1. Listeners

The subset consisted of five listeners (three female) between 23 and 30 years old (mean 26.8 years) from experiment 1, namely, S06, S07, S08, S09, and S10.

2. Stimuli and procedure

The listeners in experiment 2 performed 32 conditions of FM detection with AM masking supplementary to those of experiment 1. These extra conditions were intermingled in a randomized order with the FM detection with AM masking conditions from experiment 1, including in the initial training, in order to minimize practice effects. These extra conditions had the same parameters as those in experiment 1, except that m in Eq. (3) was fixed to -12.04 dB (re. 100% modulation; i.e., 25%) rather than -6.02 dB, and only a single stimulus duration was tested (1 s) because in experiment 1 the masking was greater for 1 than for 0.5 s. This was mostly because the baseline (i.e., FM detection without AM imposed on the carrier) FMDTs were lower for 1 than 0.5 s. The procedure was identical to that described for experiment 1.

B. Results

Figure 3 shows the effect of AM masker depth on the masking patterns for FM detection. The panels and their ordinates follow those of Fig. 2. The masking patterns for 50% masker AM depth are the means for the five subjects who completed experiment 2, so they are slightly different from the means in Fig. 2, but allow for a comparison of masking patterns across masker AM depth that is balanced across subjects.

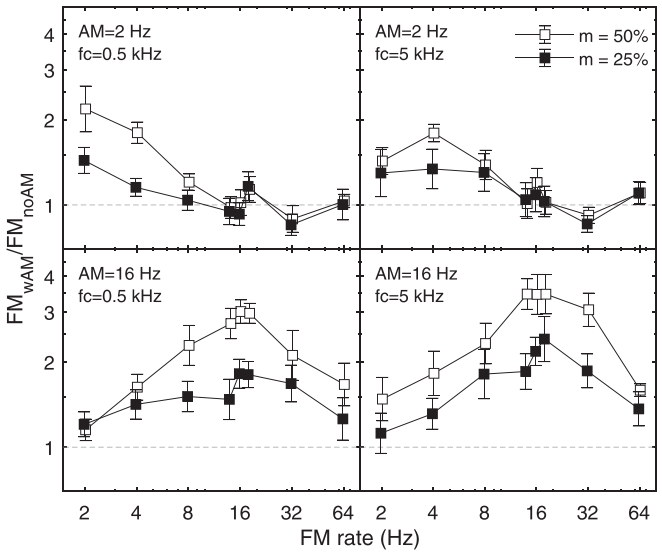

FIG. 3. Masking patterns by FM rate for an AM masker depth of 50% (open squares) and 25% (filled squares). Stimulus duration was 1 s. Panel layout is as in Fig. 2 with ±1 SE bars.

The masking patterns for 25% masker AM depth clearly show less masking than those for 50% masker AM depth. However, they do exhibit similar form. For 16-Hz masker AM, the peak of the masking function is approximately 16 Hz. For 2 Hz, the masking for 25% masker AM depth is minimal throughout as a function of FM rate. A four-way ANOVA was performed on the log-transformed masking ratios (FM_{wAM} / FM_{noAM}) with factors of FM rate (2, 4, 8, 14, 16, 18, 32, or 64 Hz), AM rate (2 or 16 Hz), AM-masker depth (25 or 50%) and f_c (0.5 or 5 kHz). It showed a significant effects of masker depth [$F(1,4) = 95.48$, $p < 0.01$, $\eta p^2 = 0.96$] and AM rate [$F(1,4) = 47.98$, $p < 0.01$, $\eta p^2 = 0.92$] and a significant interaction between the two [$F(1,4) = 94.62$, $p < 0.01$, $\eta p^2 = 0.96$]. It appears that there is a greater difference in masking ratios between 25 and 50% masker modulation depths for the 16-Hz masker than the 2-Hz masker. There was no significant effect of f_c nor FM rate. However, FM rate interacted with AM rate [$F(7,28) = 35.26$, $p < 0.001$, $\eta p^2 = 0.90$] and masker depth [$F(7,28) = 2.55$, $p < 0.05$, $\eta p^2 = 0.39$], although the latter interaction was not significant after Greenhouse-Geisser epsilon correction for lack of sphericity ($p = 0.12$). This was also true for the three-way interaction between f_c, AM rate and FM rate [$F(7,28) = 3.11$, $p < 0.05$, $\eta p^2 = 0.44$; G-G epsilon correction $p = 0.08$]. The two significant interactions (masker depth by AM rate and FM rate by AM rate) were marginal to the significant three-way interaction between masker depth, AM rate and FM rate [$F(7,28) = 8.62$, $p < 0.001$, $\eta p^2 = 0.68$]. Figure 3 suggests that the effect of masker depth varies in strength across FM rates, and is greatest when the FM rate is at or close to the AM masker rate, as hypothesized. Whilst this appears to be the case more for the 0.5-kHz f_c than the 5-kHz f_c, the four-way interaction was not significant [$F(7,28) = 0.98$, $p = 0.46$, $\eta p^2 = 0.20$]. However, with a sample of only five listeners, there was likely a lack of sufficient power for testing a four-way interaction (observed power = 0.345).

In order to compare the effect of 25%-depth and 50%-depth AM maskers (at both 2 and 16 Hz) on FMDTs against FMDTs without an AM masker—which creates a nested experimental design—the data were split by AM rate for analysis of AM-masker depth against the no-AM-masker conditions. No corrections were made for the increase in the chance of a type-I error. One ANOVA was performed for the 2-Hz AM masker conditions and another for the 16-Hz AM masker conditions, both with three factors: AM-masker depth (two levels: 25 and 50%), FM rate (eight levels: 2, 4, 8, 14, 16, 18, 32, and 64) and f_c (two levels: 0.5 and 5 kHz). The main effect of masker depth was significant for both AM rates [For 2 Hz: $F(2,8) = 23.91$, $p < 0.01$, $\eta p^2 = 0.85$; and for 16 Hz: $F(2,8) = 83.53$, $p < 0.001$, $\eta p^2 = 0.95$]. For the 2-Hz AM masker, consecutive comparisons suggest that FMDTs with a 25% masker AM depth were, when averaged over FM rates and f_c, marginally significantly higher than FMDTs without an AM masker [$t(8) = 2.95$, $p = 0.033$], and that thresholds with a 50% masker AM depth were significantly higher again those with a 25% masker AM depth [$t(8) = 3.94$, $p < 0.01$]. For the 16-Hz AM masker, FMDTs with a 25% masker depth were significantly higher than those with no AM masker [$t(8) = 7.35$, $p < 0.001$], and thresholds with a 50% masker AM depth were, in turn, significantly higher again than those with a 25% masker AM depth [$t(8) = 5.53$, $p < 0.01$]. The effect of FM rate was significant for the 16-Hz AM ANOVA [$F(7,28) = 7.05$, $p < 0.001$, $\eta p^2 = 0.64$], but not for the 2-Hz AM ANOVA [$F(7,28) = 2.07$, $p = 0.24$, $\eta p^2 = 0.34$]. Although it is not valid to compare these tests across ANOVAs, an interaction between the effects of AM rate and FM rate has already been shown in experiment 1 and for the masking ratios tested above. The interactions between the effect of masker depth and FM rate were significant [For 2 Hz: $F(14,56) = 7.42$, $p < 0.001$, $\eta p^2 = 0.65$; for 16 Hz: $F(14,56) = 7.44$, $p < 0.001$, $\eta p^2 = 0.65$]. For both 2- and 16-Hz AM maskers, FMDTs increase with masker depth more at some FM rates than at others. The effect of f_c was not significant for either ANOVA [2 Hz: $F(1,4) = 14.87$, $p = 0.07$, $\eta p^2 = 0.79$; 16 Hz: $F(1,4) = 7.79$, $p = 0.15$, $\eta p^2 = 0.66$], but for both ANOVAs the effect of f_c interacted with the effect of FM rate [for 2 Hz: $F(7,28) = 4.39$, $p < 0.05$, $\eta p^2 = 0.52$; for 16 Hz: $F(7,28) = 4.19$, $p < 0.05$, $\eta p^2 = 0.51$]. The interactions between the effect of f_c and masker depth and the three-way interactions were not significant for either ANOVA.

To summarize, consistent with our initial expectations, increasing the depth of the AM masker increased the masking between AM and FM. Surprisingly, an effect of AM depth was observed even in conditions in which FM is not assumed to be encoded by a temporal-envelope code.

IV. EXPERIMENT 3: EFFECTS OF FM-AM PHASE RELATIONSHIP

FMDTs with AM masking were measured as a function of the phase difference between the FM and AM. Following the design of Strickland and Viemeister (1996), thresholds were tested with the AM masker modulation rate at half the target FM rate, the same rate, and twice the target FM rate.

Due to timing constraints, the design was not fully factorial. Instead, listeners were tested for phase effects on masking in two combinations of FM rate and f_c only, namely, a 2-Hz FM and 0.5 kHz f_c (for which FM is assumed to be encoded by a temporal-fine-structure code) *versus* a 16-Hz FM and 5 kHz f_c (for which FM is assumed to be encoded by a temporal-envelope code). However, four of the six listeners were also tested with a combination of 2-Hz FM and 5 kHz f_c for comparison between the carrier frequencies at a slow FM rate on one hand, and between fast and slow FM rates for a high f_c on the other.

A. Methods

1. Listeners

Another subset of six listeners (five female) between 20 and 28 years old (mean 23.8 years) from experiment 1, namely, S01, S02, S03, S04, S05, and S06, completed experiment 3.

2. Stimuli and procedure

The stimuli were created according to Eqs. (2) and (3) in Sec. III A 2. However, the starting phase of the FM [Φ_{FM} in Eq. (2)] was either 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. The AM starting phase [Φ_{AM} in Eq. (3)] was always 0°. For example, for equal FM and AM rates, an FM starting phase of 180° would produce an upper frequency excursion limit ($f_c + \Delta f$) at the trough of the AM cycle and a lower frequency excursion limit ($f_c \Delta f$) at the peak of the AM cycle. Again, f_c was either 0.5 or 5 kHz with a starting phase φ varied randomly [$U(0,2\pi)$] on every stimulus. When f_c was 0.5 kHz, the FM rate was always 2 Hz and the AM rate was either 2 or 4 Hz. When the f_c was 5 kHz, the FM rate was always 16 Hz and the AM rate was 8, 16, or 32 Hz. Therefore, f_c and FM rate co-varied and were not analyzed separately. The tracking procedure and step sizes were the same as those described for FM detection with an AM masker in experiments 1 and 2. The order of conditions was randomized. However, listeners S01, S04, S05, and S06 also completed further conditions with an f_c of 5 kHz, FM rate of 2 Hz and AM rates of 2 and 4 Hz in a separate testing session at a later date. Masker m was fixed at −6.02 dB and stimulus duration was fixed at 1 s.

B. Results

The FMDTs as frequency excursion proportional to f_c are plotted in Fig. 4. There was a trend for more masking at 90° (FM re. AM) than at 0°, 225°, 270°, and 315° when both FM and AM rate = 16 Hz and $f_c = 5$ kHz. In the remaining conditions, there was little effect of phase if any on the masking of AM on FM detection. A three-way ANOVA with factors of phase (eight levels), FM rate and f_c (two levels: 2-Hz FM with 0.5 kHz f_c, and 16-Hz with 5 kHz f_c) and AM rate (two levels: equal to f_{FM}, and double f_{FM}) was performed on the log-transformed FMDTs. The main effects of phase, FM rate and AM rate were not significant. The two-way interactions between phase and FM rate and between phase and AM rate were not significant. The two-way

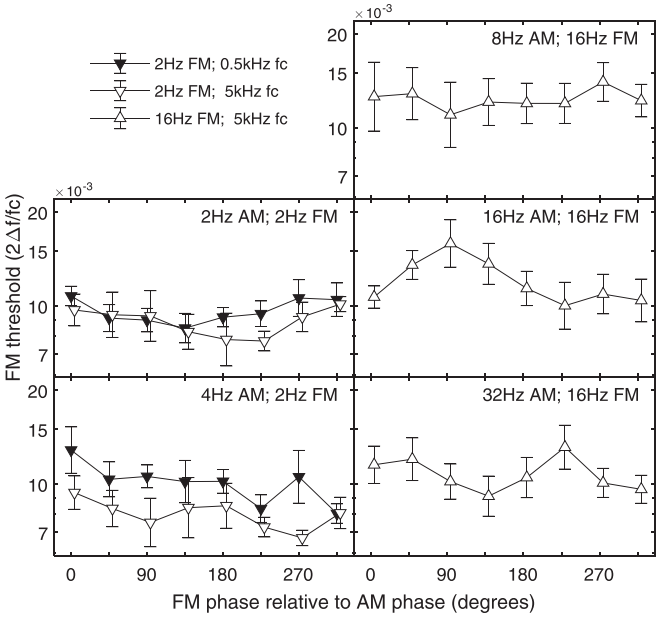

FIG. 4. FMDTs with an AM masker ($m = 50\%$) at half (top row), the same (middle row), or double (bottom row) the FM rate, as a function of FM phase. AM was fixed at sine phase. Upwards and downwards pointing triangles denote an FM rate of 2 and 16 Hz, respectively. Filled and open symbols denote an f_c of 0.5 and 5 kHz, respectively. Stimulus duration was 1 s. Error bars = ±1 SE.

interaction between FM and AM rate was significant [$F(1,5) = 11.48$, $p < 0.05$, $\eta p^2 = 0.70$]. For the 2-Hz FM rate (and $f_c = 0.5$ kHz), thresholds were generally lower with an AM rate of 2 Hz than of 4 Hz, whereas the for 16-Hz FM rate (and $f_c = 5$ kHz), thresholds were generally higher with an AM rate of 16 Hz than of 32 Hz. The three-way interaction between phase, FM rate and AM rate was significant [$F(7,35) = 4.6$, $p < 0.001$, $\eta p^2 = 0.48$). To interpret this, two separate two-way ANOVAs were conducted: one for $f_{FM} = 2$ Hz (or $f_c = 0.5$ kHz) with factors of phase (8 levels) and AM rate (2 levels: 2 and 4 Hz), and one for $f_{FM} = 16$ Hz (or $f_c = 5$ kHz) with factors of phase (8 levels) and AM rate (3 levels: 8, 16, and 32 Hz). The main effects of phase and AM rate were not significant for either ANOVA. However, the interaction between phase and AM rate was significant for the ANOVA with $f_{FM} = 16$ Hz [$F(14,70) = 2.66$, $p < 0.01$, $\eta p^2 = 0.35$], whilst it was not significant when $f_{FM} = 2$ Hz [$F(7,35) = 1.65$, $p = 0.15$, $\eta p^2 = 0.25$].

The supplementary conditions performed S01, S04, S05, and S06 allowed for comparison between f_c at 2-Hz FM and between FM rates at $f_c = 5$ kHz. At 2-Hz FM and AM, neither an f_c of 0.5 or 5 kHz shows any phase effects and FMDTs are generally of the same magnitude. At 4-Hz AM and 2-Hz FM, thresholds are lower with an f_c of 5 kHz than an f_c of 0.5 kHz. However, an ANOVA of FMDTs for 2-Hz FM with factors for phase, AM (2 or 4 Hz) and f_c (0.5 or 5 kHz) showed no significant effects or interactions; only the main effect of phase trended towards significance [$F(7,21) = 2.22$, $p = 0.07$, $\eta p^2 = 0.43$], but no trend is clearly visible in Fig. 4.

At an f_c of 5 kHz, FMDTs were lower and showed less effect of phase for 2-Hz FM than for 16-Hz FM. However, an ANOVA of the thresholds when $f_c = 5$ kHz with factors

of phase, FM (2 vs 16 Hz) and AM (equal or twice FM rate) did not reveal any significant effect of FM or interaction between FM and phase. Only a significant interaction between phase and AM was found [$F(7,21) = 2.79$, $p < 0.05$, $\eta p^2 = 0.48$], suggesting the change in threshold across phase was different when AM rate was twice FM rate compared to when the rates were equal. In Fig. 4, this is clear for 16-Hz FM, but not 2-Hz FM. With only four listeners, however, these further analyses are likely to be underpowered.

V. MODEL SIMULATIONS

As the behavioral results described above show that AM degrades FM detection with the features of broad tuning around the masker rate and an effect of masker depth, it is possible that the masking patterns can be explained by AM processing that considers only the temporal dynamics of excitation patterns (that is, temporal-envelope cues), be it AM externally applied to a signal or AM created by FM after cochlear filtering. A simple model of a system that detects modulations in amplitude was tested to determine if it could reproduce the FMDTs and the masking produced by AM.

A. Model specification

The model structure is similar to that used by Wallaert *et al.* (2017). Three implementations of the model were tested with respect to the role of envelope phase at the output of the modulation filters (see stage VII in the list below). This was motivated by the fact that the exact modulation rate above which listeners lose envelope phase sensitivity ranges between 6 Hz (Dau, 1996, p. 35) and 12 Hz or even one octave higher (Sheft and Yost, 2007a). The first implementation retained the envelope phase by passing the outputs of the modulation filter-bank (stage VI) directly to the noise-addition stage (the "phase-preserved" model). The second implementation discarded the envelope phase of all the modulation filter-bank channels by passing only the absolute magnitude of the Hilbert transform of the outputs to the noise stage (the "phase-discarded" model). The third implementation progressively reduced sensitivity to envelope phase with increasing modulation filter center-frequency by passing the modulation filter output directly for channels below 6 Hz and passing the half-wave rectified, low-pass filtered (2nd order Butterworth with 6 Hz cut-off) output above 6 Hz (the "phase-reduced" model).

In addition, each implementation of the model had two sub-variants for the conditions of FM detection in the presence of an AM masker: one that used a template (see stage IX below) constructed using the AM masker in the signals (thus giving the model beating cues to detect the target resulting from the interaction between the AM and FM) and one that used a template constructed without the AM masker in the signals (thus ignoring any AM-FM beat cues). The model had the following stages in sequential order:

(i) a bank of five gammatone filters, one centered at the f_c of the stimulus, and the remaining four centered at 1 and 2 ERB_N above and below the f_c of the stimulus;

(ii) a "broken-stick" input-output function for the output of the gammatone filter tuned to the f_c of the stimulus; the function is linear up to a knee-point of 30 dB SPL and compressive (using a power law with an exponent of 0.3) above;

(iii) half-wave rectification of all five channels;

(iv) high-pass filtering (1st order 3 dB/oct roll-off, 3-Hz cut-off) of all channels to simulate short-term adaptation (Tchorz and Kollmeier, 1999);

(v) the onsets and offsets were removed by cutting the first 100 ms and the final 20 ms of each channel. This was done to force the modulation filter-bank to process only the ongoing modulations and "ignore" the interval onsets and offsets;

(vi) the signal of each channel was passed to a filter-bank (1st order Butterworth filters) with ten logarithmically-spaced channels between 2 and 120 Hz (Moore *et al.*, 2009), each with a Q factor of 1 (Ewert and Dau, 2000) to decompose the modulations of the processed signals, producing 50 channels;

(vii) one of the three treatments described above is performed on the output of the modulation filters: either retaining the envelope phase information, discarding it from all modulation filters, or progressively reducing it with increasing filter center-frequency;

(viii) three independent Gaussian noises (an "additive," a "multiplicative," and a "memory" noise) were added consecutively to the output of all 50 channels; the first type of noise ("additive") had a constant standard deviation (SD) (Dau *et al.*, 1997); the second type of noise ("multiplicative") had an SD proportional to the rms of each channel (SD = 1 dB re: rms; Ewert and Dau, 2004; Wallaert *et al.*, 2017); the third type of noise ("memory") was additive like the first one, but had an SD which was multiplied by an exponential decay function to model echoic-memory limitation; the addition of this "memory noise" resulted in a weaker representation of the earlier part of the signal than the later and reduced temporal integration of envelope cues (Ardoint *et al.*, 2008; Wallaert *et al.*, 2017), affecting the representation of longer duration stimuli more than shorter duration stimuli. The decay time constant was fixed at 1.2 s;

(ix) The final decision stage was based on a template matching process (Dau *et al.*, 1997). The model created a template at the start of each staircase with the DV set at the starting value and without any added noise. The template was calculated as the difference between the internal representations of the target and reference stimuli (with or without the AM masker for the FM detection staircases with an AM masker). On each trial, the target and reference stimulus intervals were cross-correlated channel by channel with the template and divided by the product of the rms of the two signals (i.e., normalized). The lags used in the cross-correlation were restricted to $\pm$ 1 target modulation cycle. The interval with the largest normalized cross-correlation coefficient (summed across channels) was selected by the model.

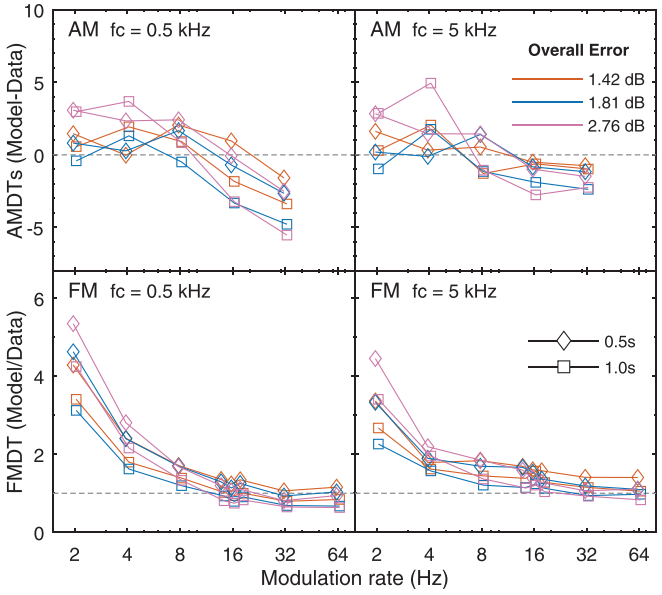

FIG. 5. (Color online) Differences between simulated and real AMDTs (top) and ratios between simulated and real FMDTs (bottom) by modulation rate for the three model variants: preserving the phase of the modulation filter outputs (orange), discarding the output phase of all the modulation filters (blue), and preserving the phase below 6 Hz and being progressively less sensitive to it above 6 Hz (pink). Diamonds and squares denote 0.5 and 1.0 s stimuli, respectively.

Each implementation of the model was fitted to predict the AMDTs by adjusting the SD of the "additive noise" and the "memory noise." The best fits yielded overall rms errors (the difference in thresholds between simulated and real data) of less than 3 dB for all three implementations (see top panels of Fig. 5); the best overall goodness-of-fit was for the "phase-preserved" model implementation (1.42 dB) and the worst was for the "phase-reduced" implementation (2.76 dB). With these noise values, the models performed the FM detection task with and without AM maskers, as the listeners did. The only difference in procedure was the number of reversals required before terminating a staircase. For the models this was 16 rather than 8. The models performed 50 staircases for each condition to achieve a reliable threshold estimate.

B. Simulated results

1. AM detection thresholds

In the top panels of Fig. 6, the behavioral results are overlaid with the model simulations for the three implementations with the rms error for each duration and f_c combination inset. These plots and those in the top panels of Fig. 5 show that the model simulations fit the behavioral data relatively well (particularly for the 0.5-s duration stimuli and the 5-kHz f_c). However, for the 1-s duration stimuli and 0.5-kHz f_c, the predicted thresholds decrease more rapidly than the real thresholds when modulation rate increases (overestimating thresholds at 2 and 4 Hz and underestimating thresholds at 16 and 32 Hz). The "phase-preserved" model (leftmost set of panels in Fig. 6) matches the behavioral data best. The "phase-discarded" model (middle set of panels in Fig. 6) produced a more linear decrease in AMDT with AM rate (thus a less good fit to the data), but the "phase-reduced" (rightmost set of panels in Fig. 6) produced the most linear and poorest predictions of AMDTs.

2. FM detection thresholds

For FM detection without an AM masker, all three versions of the model only predicted FMDTs for FM rates of 8 Hz and above (see the bottom panels of Figs. 5 and 6). The bottom panels of Fig. 5 show the ratio between the modelled and behavioral FM results. For 2 and 4 Hz, the models performed poorer than the human listeners, overestimating thresholds by a large amount (factor from 3 to 5 at 2 Hz). Contrary to the AM simulations, the FM simulations fitted the behavioral data better for the 1.0-s stimuli than the 0.5-s

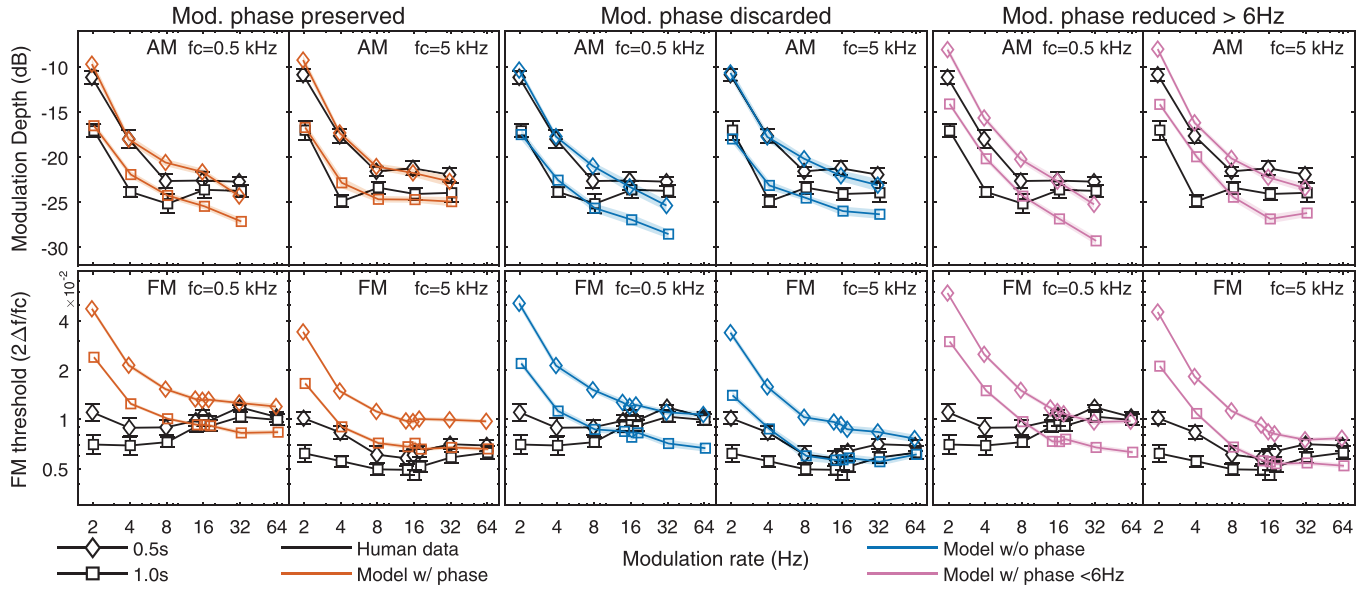

FIG. 6. (Color online) AMDTs and FMDTs (as in Fig. 1) overlaid with simulations by the three variants of the modulation filter-bank model: modulation filter phase preserved (left set of panels, orange), modulation filter phase discarded (middle set of panels, blue), and modulation filter phase sensitivity decreasing above 6 Hz (right set of panels, pink). The shaded areas around the simulated means denote ±1 SE.

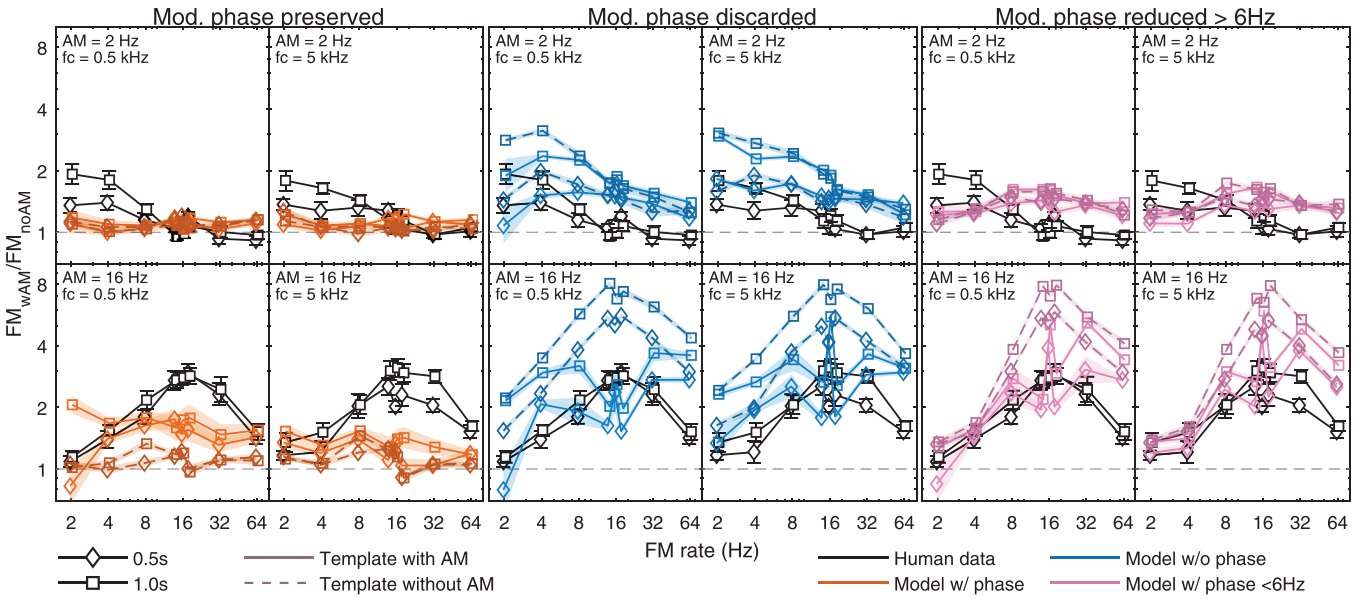

FIG. 7. (Color online) AM-masking-FM patterns by FM rate for 0.5 and 1.0 s stimulus durations, as in Fig. 2, with the simulations of the three model variants overlaid in color (as in Fig. 6). The solid and dashed colored lines show the simulations from the model sub-variants that made templates with and without AM, respectively.

stimuli. However, for both AM and FM, the simulations fit the data better at 5-kHz f_c than 0.5-kHz f_c. At low FM rates, the "phase-preserved" and "phase-discarded" models perform better than the "phase-reduced" model. At high FM rates, the "phase-reduced" model fits the data bests at 5 kHz f_c, but the "phase-preserved" model fits best at 0.5 kHz f_c.

Overall, the models performed more poorly than real listeners at low modulation rates, but at higher modulation rates, the models performed better than real listeners for AM detection and relatively accurately for FM detection.

3. AM-masking-FM: Tuning and beating

The masking patterns produced by the models are plotted (overlaying the behavioral data) in Fig. 7 following the same order of the models as in Fig. 6. As in Fig. 2, The FMDTs with an AM masker ($m = 50\%$) were divided by the FMDTs without an AM masker (bottom panels of Fig. 5). For the 2 Hz masker, none of the versions of the model that were tested were able to successfully replicate the behavioral data. The "phase-preserved" model produced little to no masking. The "phase-discarded" and "phase-reduced" models produced some masking, but the masking lacked the tuning seen in the behavioral data. In fact, the "phase-reduced" model predicted *more* masking for $f_{FM} \geq 8$ Hz than $f_{FM} < 8$ Hz. The "phase-discarded" model shows more masking for lower than higher FM rates, but in general predicts too much masking, with the peak sometimes at 4 or 8 Hz, and masking spreading up well above 8 Hz (which was not observed behaviorally). Whatever the mechanism may be that caused the *tuned* masking of 2- and 4-Hz FM by 2-Hz AM in the behavioral data, it is not explicable by the processing of the temporal-envelope cues available to the models.

For the 16 Hz masker, the masking patterns produced by the models depended greatly on whether the AM was included in the creation of the template or not. This provided beating cues when the AM and FM rates were close in rate. The two sub-variants of each model: with the AM included in, or excluded from, the template is respectively displayed by solid and dashed, colored lines in Fig. 7. The "phase-discarded" and "phase-reduced" models showed large amounts of masking when the AM was excluded from the template, and it was tuned around 16 Hz. However, it was more than twice the magnitude of masking observed behaviorally. When the AM was included in the template, the amount of masking, particularly at $f_{FM} = 14$ and 18 Hz, was dramatically reduced. However, a sharp peak at $f_{FM} = 16$ Hz remained, sometimes reaching the same magnitude of masking as the models with the beat cues. This suggests that the models were able to use envelope-beat cues (second-order envelope cues) to reduce AM-masking of FM, if the models are provided with these cues. However, this is inconsistent with the behavioral data which, like the simulations from the models that ignore the envelope-beat cues, do not display the release from masking at 14 and 18 Hz.

The versions of the model that produced AM-masking-FM (the "phase-discarded" and "phase-reduced" models) were used to also produce simulations of the second and third behavioral experiments (i.e., AM-masker depth effects and FM-AM phase relation effects), both with and without the AM in the template.

4. AM-masking-FM: Masker depth effects

The simulations of the AM-masker depth effects are plotted in Fig. 8. The simulated masking patterns with an AM masker of $m = 50\%$ are copied from Fig. 7. For both models, less masking was seen when $m = 25\%$ than when $m = 50\%$, particularly for the 16-Hz AM masker. The reduction in masking is comparable to that produced when the stimulus duration was halved (from 1 to 0.5 s). It is

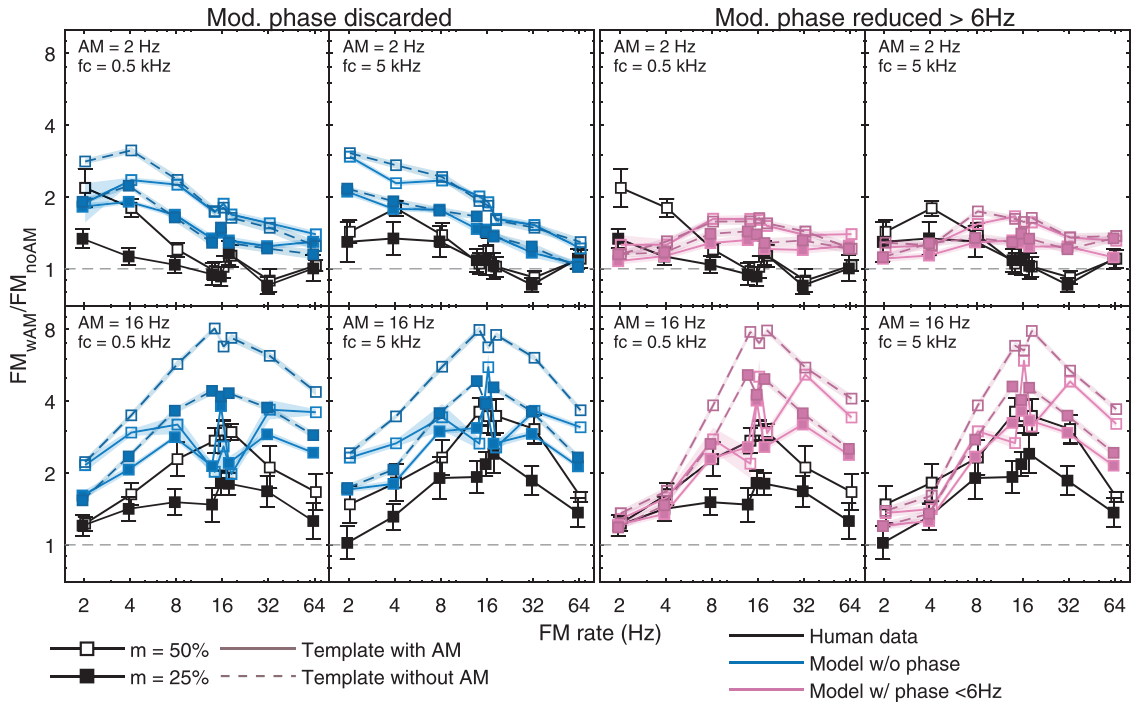

FIG. 8. (Color online) Masking patterns by FM rate for AM masker depths of 50 and 25%, as in Fig. 3, with the simulations of the models overlaid as in Figs. 6 and 7. Only the models that produced substantial *tuned* masking were tested. Namely, those that discarded the phase of all the modulation filters (left set of panels, blue) or progressively reduced sensitivity to the modulation phase above 6 Hz (right set of panels, pink). Again, solid and dashed lines denote models with and without AM in the template, respectively.

approximately the same size reduction, on a logarithmic scale, as seen in the behavioral data. However, the reduction was less than equivalent to a halving of the masking effect, as the models overestimated masking. The effect of masker depth is clear for the simulations made without AM in the template. On the other hand, the simulations with the beating cues showed less of an effect of masker depth, particularly when the FM and AM rates were equal or close. Again, notches in the masking pattern at $f_{FM} = 14$ and 18 Hz were seen for the 16-Hz AM masker conditions, indicating the usage of beat cues.

5. AM-masking-FM: Phase effects

Finally, simulations were run in the conditions of experiment 3. The results of which can be seen in Fig. 9. As before, dashed lines denote the versions of the models without the beating cue in the template and solid lines denote the versions with the beating cue in the template. As the models were much less sensitive to slow-rate FM than the human listeners, masking is plotted rather than thresholds in terms of frequency excursion as a proportion of f_c (as in Fig. 4), in order to bring the behavioral and model data into a comparable range. For each listener and each model, the FMDTs from experiment 3 were divided by the FMDTs without an AM masker (from experiment 1) at the same FM rate to give the ratios in Fig. 9. Both models ("phase-discarded" and "phase-reduced") produced similar results for 16-Hz FM: they predicted similar amounts of masking and similar differential effects of inclusion versus exclusion of AM in the template. When the AM masker was 8 Hz, neither model predicted phase effects (regardless of presence or absence of AM in the template), which is in line with the behavioral data. Only the "phase-discarded" model with a template with AM predicted roughly the correct masking magnitude; the others overestimated it. When the AM masker was 16 and 32 Hz, both models predicted bimodal patterns across phase. When the AM masker was 16 Hz, these patterns were in anti-phase for templates with and without AM; using a template with AM produced very strong phase effects (ranging from ratios of 1 to 5 or 6) with peaks at 0 and 180°, whereas a template without AM produced peaks at 90 and 270° and a phase effect of similar size to (on a log scale) the behavioral data, but overestimated overall masking magnitude. For the 32-Hz AM masker, the bimodal patterns produced with a template with AM were in phase with, and of a similar effect size to, those observed behaviorally (peaks at 45 and 225°), but overall masking was overestimated. Phase effects produced with a template without AM were also in phase the behavioral data, but overestimated in both effect size and overall masking magnitude.

For the 2- and 4-Hz AM maskers (i.e., for a 2-Hz FM), the models behaved quite differently. Generally, the "phase-reduced" model underestimated masking, but produced similar phase patterns whether the template was made with or without the AM. For the 2-Hz AM, it predicted no masking overall and only slight variations across phase. With 4-Hz AM, it predicted peaks in masking at 45 and 225°, with more pronounced peaks without AM in the template. However, the behavioral data did not show any masking effects with a 2-Hz FM. The "phase-discarded" models with and without AM in the template produced bimodal phase patterns in anti-phase with each other, toughing and peaking at 90 and 180°, respectively, for 2-Hz AM and 45 and 225°

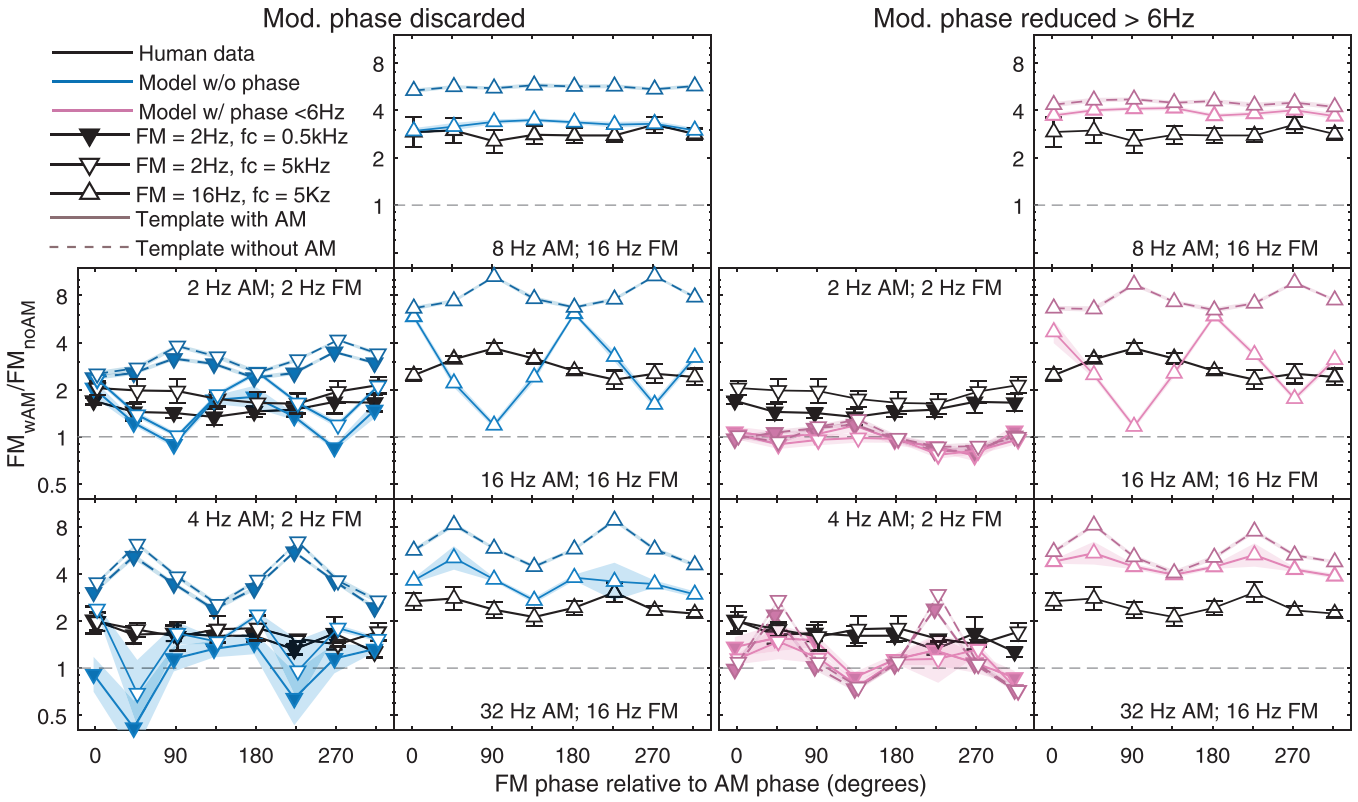

FIG. 9. (Color online) Masking patterns as a function of phase difference between FM and AM. Masking ratios calculated using FMDTs without an AM masker from experiment 1. Panel order and symbols follow Fig. 4. The simulations from the models that discarded modulation phase in all filters (left, blue) or filters above 6 Hz (right, pink) are overlaid in colored solid or dashed lines for with and without AM in the template, respectively.

for 4-Hz AM. Using a template with AM generally underestimated the masking effect, whilst without AM overestimated it.

Overall, the simulations of the effect of the phase difference between the FM and AM on masking showed that the models consistently predicted bimodal patterns across phase. These patterns were sometimes in phase with, and of a similar magnitude, to the behavioral data (i.e., at 32-Hz AM and 16-Hz FM), but otherwise they overestimated phase effects and strong bimodal phase effects were also predicted when little effect of phase was seen behaviorally (such as for the 2-Hz FM). The "phase-reduced" model was better than the "phase-discarded" model at predicting no effect of phase in the conditions where little effect was seen behaviorally.

VI. GENERAL DISCUSSION

The current study demonstrates that, under certain circumstances, AM impairs the detection of FM for normal-hearing listeners. This has been shown previously (Moore and Sek, 1996; Moore and Skrodzka, 2002; Ernst and Moore, 2010; Paraouty *et al.*, 2016; Paraouty and Lorenzi, 2017). The current study contributes (i) in-depth characterization of how the masking of FM by AM changes as a function of FM and AM rate, AM depth, f_c, stimulus duration and phase difference between AM and FM, and (ii) insights into possible role of temporal-envelope processing mechanisms through computational modelling based on the modulation filter-bank concept.

A. Main findings

The first two findings were: (i) the observation of masking of FM for both slow and fast AM-masker rates and low and high carrier frequencies; and (ii) for all conditions the masking of FM by AM was broadly tuned and was frequency selective (most masking occurred at the masker rate). This is consistent with the masking patterns found for AM-masking-AM in previous psychophysical studies and estimates of frequency selectivity in the temporal-envelope domain (Houtgast, 1989; Bacon and Grantham, 1989; Strickland and Viemeister, 1996; Ewert and Dau, 2000; Lorenzi *et al.*, 2001; Sek *et al.*, 2015).

The third finding was the absence of beating effects between AM and FM when AM and FM rates are close, in contrast with the beating effects previously reported for AM-masking-AM experiments and complex-AM discrimination tasks (Strickland and Viemeister, 1996; Lorenzi *et al.*, 2001; Millman *et al.*, 2002).

The fourth finding is that the depth of the AM masker had an effect on masking of FM. A shallower modulation depth produced less masking. This is consistent with AM-masking-AM data in previous psychophysical studies (Bacon and Grantham, 1989; Strickland and Viemeister, 1996). Paraouty *et al.* (2016) found a doubling of AM-masking-FM for a change in masker depth from 33 to 66% with a carrier of 0.5 kHz and FM and AM rates of 5 Hz. The current findings show an approximately similar doubling of masking for a doubling of masker depth from 25 to 50%, when FM and AM rates are equal.

The fifth finding is that the phase of the FM relative to the phase of the AM masker had an effect on the amount of masking for fast FM and AM rates (16 Hz) and a high f_c (5 kHz), but no effect on masking for slow FM and AM rates (2 Hz), regardless of f_c. This is somewhat consistent with Strickland and Viemeister (1996), who found only weak, variable phase effects at 2- and 4-Hz AM maskers (4 Hz target). However, they found no phase effects for target and masker AM = 16 Hz, but instead a single peak at 270° for a masker of 8 Hz (target 16 Hz). One limitation here is that Strickland and Viemeister (1996) and Bacon and Grantham (1989) used noise carriers, whilst the current study used pure-tone carriers; the impact of carrier type on modulation-masking phase effects is still unknown.

In summary, masking between AM and FM was found for both carrier frequencies and both AM-masker rates. The masking exhibited two features of AM-masking-AM—namely, broad tuning and depth effects. However, beating effects were not found and phase effects were found only at an f_c of 5 kHz for an AM rate of 16 Hz.

Computational models using only temporal-envelope cues extracted at the output of a modulation filter-bank were fitted to the AM data and then used to simulate FM detection in all tested conditions. Below 16 Hz, these models predicted much worse unmasked FMDTs than the listeners produced. Predictions of masking effects depended on whether the model preserved or discarded the phase information from the modulation filter outputs; if the phase information was preserved, masking effects were effectively not predicted. If the phase information was discarded from all modulation filter channels, masking was over-predicted, but with roughly the correct tuning. If the phase information was progressively reduced with increasing filter center-frequency, then the correct amount of masking was predicted for a 2-Hz masker, but mistuned to higher FM rates. Again, masking was over-predicted for a 16-Hz masker. The latter two models predicted duration, beating, phase and masker-depth effects, although only duration and masker-depth effects were observed behaviorally. The prediction of beating and phase effects was expected from these models as the envelope patterns elicited by the AM and FM will sum constructively and destructively in a cyclic manner depending on the frequency or phase difference, resulting in temporal-envelope cues.

B. Mechanisms responsible for FM detection

Altogether, these results are broadly consistent with the idea that, irrespective of f_c, fast (>4–8 Hz) FM is detected via temporal-envelope cues resulting from FM-to-AM conversion at the output of cochlear filters. However, whilst the computational model using temporal-envelope cues, but discarding the phase of these envelope cues, predicted the general shape of masking patterns for both slow and fast AM maskers, the magnitude of masking was over-predicted. This suggests that FM detection in human listeners is somewhat resistant to interference from masking AM. This might be because FM evokes temporal envelopes of more varied phase across cochlear-filtering channels than AM does (i.e., in anti-phase in channels above and below the f_c), and possibly the auditory system detects FM by comparing the phase of the envelopes across cochlear "channels" (Moore *et al.*, 2018). The "phase-preserved" model had access to this envelope phase information for all 50 channels (ten modulation filters by five gammatone filters) and it predicted no masking for a 2-Hz AM masker and only weak, poorly tuned masking for a 16-Hz masker. The models did not strictly compare information across channels, but rather cross-correlated each modulation-by-gammatone channel output for the test stimuli with the corresponding channel output in the template (without internal noise). It is possible that, (a) humans do not perfectly preserve a template (or internal representation) of the target in memory and/or the template is noisy; and/or that (b) humans do not compare the test stimuli with the internal template across all their cochlear-filtering channels. It is also possible that the cross-correlation comparison in the models is more efficient or somehow more resistant to noise than the "true" mechanism in humans. Essentially, the models might have over-performed on the comparison and decision stage.

1. Lack of contribution of envelope-beat cues

The combined psychophysical and modelling data suggest that, although beating envelope cues resulting from the interaction between FM and AM at the output of the cochlear filters might aid detection of FM masked by AM, listeners ignore these cues. This contrasts with AM-masking-AM data (e.g., Strickland and Viemeister, 1996; Millman *et al.*, 2002). This suggests that real listeners adopt a sub-optimal listening strategy: they may not build up a template of the target by comparing internal representations of the target and comparison intervals, utilizing the second-order envelope cues in the target interval, but rather they construct an internal representation of the target FM alone, independent of its interactions with competing sounds (e.g., AM maskers). Why would they do this despite the fact that, according to listeners' anecdotal reports, these envelope beats are salient at supra-threshold FM excursions? One possibility might be that the envelope beat is perceived as a slow loudness fluctuation. This does not correspond to the pitch cue that the listeners were trained to listen for in all the other conditions where beating between the FM and AM did not occur. Thus, the experimental context probably plays a greater role in building the internal template than expected from the model (which performs every staircase independently).

2. Two mechanisms of FM detection

In the case of the 2-Hz AM masker, masking of the correct magnitude and tuning could not be predicted by the computational model using only temporal-envelope cues extracted at the output of a modulation filter-bank. This suggests that for slow FM rates (≲5 Hz), FM is not detected via temporal-envelope cues resulting from FM-to-AM conversion. This is also consistent with the fact that phase effects between AM and FM were found for fast FM and AM rates and a high f_c (5 kHz), but not for slow FM and AM rates regardless of f_c (0.5 or 5 kHz).

Whilst masking was significantly smaller for the 2-Hz AM masker compared to the 16-Hz AM masker, the lack of a significant effect of f_c on masking does not support the hypothesis that TFS cues are used to detect slow-rate FM at low carrier frequencies. If this were the case, one would expect much less masking (if any) for the low f_c (where neural phase locking is most accurate; Palmer and Russell, 1986) than for the high f_c, at slow FM and AM rates. This was proposed by Moore and Sek (1996), based on masking of FM by AM ($m = 33\%$) at 2 Hz being greater at 4 and 6 kHz than 0.5 kHz. However, they did find some masking even at slow modulation rates and a 0.5 kHz carrier, and it was of a similar magnitude to that found in the current study for the 2-Hz FM and AM and an m of 25%. Moore and Sek (1996) found no effect of modulation rate when $f_c = 6$ kHz, but substantially more masking at 10 and 20 Hz than 2 Hz modulation rates when $f_c = 0.5$ kHz, whereas the current study found a similar increase in masking from 2 to 16 Hz modulation rates ($f_{FM} = f_{AM}$) for both carrier frequencies. More recently, Moore *et al.* (2018) demonstrated better discrimination of AM and FM at a rate of 2 Hz than at a rate of 10 Hz, when AM and FM are equated in detectability. They claim this as evidence of TFS cues aiding slow-rate FM processing, and suggest that the hypothesis that phase of the modulation across the excitation pattern is compared to discriminate AM (in-phase) from FM (producing AM in anti-phase) is not sufficient to explain further, preliminary (but not reported) data.

It is not possible to rule out phase-locking as the encoding mechanism of low-rate FM at low carrier frequencies, but if this is the case, it is also affected by an interfering AM. Significant level effects on pitch discrimination have been demonstrated for pure-tone stimuli (Emmerich *et al.*, 1989). It is possible that changes in level due to the AM masker 2-Hz cycle produced variations in pitch which in turn produced the AM-masking-FM seen at slow FM rates in the current study (at an f_c of 0.5 kHz with a 2-Hz AM masker). However, the changes in pitch due to stimulus level found by Emmerich *et al.* (1989) are weak and vary across listeners; at 0.5 kHz, most listeners do not report hearing any changes in pitch, whereas the masking effect observed in the current study (when $f_c = 0.5$ kHz with a 2-Hz AM masker) was relatively robust and consistent across listeners. Therefore, it is unlikely that this masking reflects the detrimental effect of level variations on the operation of a pitch mechanism using excitation-pattern (i.e., place) cues.

3. Post-sensory explanations of masking effects

The origin of the masking effect at slow modulation rates for both low and high carrier frequencies might be cognitive rather than purely sensory. According to this interpretation, this masking would occur at a late (post-sensory) stage of auditory processing, where the available sensory information conveyed by FM has been transformed into a single fluctuating pitch percept, regardless of encoding mechanism (be it a single mechanism such as place-rate information, or by separate mechanisms at high and low audio frequencies). The listeners might then be confused by the competing temporal modulation of the masker at the same or similar rate of the target FM, even though the two modulations elicit different percepts (loudness versus pitch fluctuations). This confusion would possibly result from grouping of the modulations into a single auditory object due to the similarity of rate. This interpretation would then correspond to a case of informational masking. Alternatively, the masking could be interpreted as a result of the AM making it more difficult to attend to the FM. That is, the two modulators compete for attention. This would also be a case of informational masking. Such an argument has been made to explain modulation masking when the modulators act on carriers of different frequencies (Sheft and Yost, 2007b). Sheft and Yost (2007b) found that masking did not follow predictions of energetic masking or grouping due to similarity of rate in the modulation domain. The failure of the models to capture the tuning of masking at slow modulation rates based on purely temporal-envelope cues is consistent with the idea that at least some of the AM masking effects might be related to informational masking. It is important to note that these interpretations of masking effects for slow AM and FM rates do not preclude a role of a TFS code in FM detection. Still, further work is warranted to demonstrate whether the interference effect does, indeed, correspond to informational masking at a late stage of auditory processing.

C. Contribution to auditory modelling of FM perception

1. Modelling FM detection

Overall, the modelling section of the study suggests that the perceptual model of modulation processing developed by Dau and colleagues can account for FM detection at rates above 8 Hz, but it does not predict the masking effect of an interfering AM unless it disregards the envelope phase at the output of the modulation filters (in which case it overestimates the masking effect). If envelope-beat cues are ignored (by exclusion of the AM for the model's template), the predicted masking is appropriately tuned, which is not surprising considering that the model predicts AM-masking-AM relatively well (Ewert and Dau, 2000). A model with a compensatory mechanism that reduces the effect of an uninformative (non-target) AM may better explain the behavioral data at both slow and fast FM rates. A dual-path model combining TFS and temporal-envelope cues (e.g., Ewert *et al.*, 2018) may better account for FM detection at slow FM rates, but it is unlikely to reproduce the masking seen at slow modulation rates. Furthermore, it would not better predict the magnitude of masking at fast modulation rates as a TFS pathway is generally considered too "sluggish" to capture information about fast FM rates (e.g., Moore and Sek, 1996).

2. Relevance to speech perception

The current results indicate that the AM and FM features of competing speech sounds should interfere, but the features (i.e., magnitude and dynamics such as phase effects) of this interference differ depending on whether the AM and FM are above, or below, approximately 10 Hz. This should be taken into account when exploring speech perception against concurrent speech sounds for normal hearing

listeners, but also for different groups of listeners such as hearing-impaired listeners and cochlear implantees (e.g., Stickney *et al.*, 2004; Stickney *et al.*, 2005; Zeng *et al.*, 2005).

VII. CONCLUSIONS

The current study aimed to assess if AM-masking-FM could fit into the concept of modulation filters, or whether a separate mechanism (possibly using TFS cues) was needed to explain slow-rate FM at low carrier frequencies. We found that for normal-hearing listeners, the masking of FM by AM was broadly tuned, increased with stimulus duration and with masker depth. Phase differences between the FM and AM appeared to affect masking for fast AM and FM and a high f_c, but not for slow AM and FM with either a low or high f_c. Listeners did not seem to benefit from envelope-beating cues between the FM and AM (i.e., a release from masking was not seen) when the two modulations were close in rate. Computational models implementing the modulation filter-bank concept and a template-matching decision strategy (that is, computational models using temporal-envelope cues only) suggested that the tuning of masking FM by AM at faster rates could be explained by a lack of sensitivity to the envelope phase at the output of the modulation filters, but the magnitude of the masking was overestimated. The models also suggested that, if envelope-beating cues are used when building an internal representation of the target sound, a release from masking *should* be observed, and that only by building an internal representation based on the FM alone does the model predict the behavioral data. The fact that listeners adopt a sub-optimal strategy and ignore the envelope beats suggest that the experimental context should be taken into account when modelling modulation perception.

Although AM-masking-FM was greater at fast than slow masker rates, it was still significant for a 2-Hz AM masker for both an f_c of 0.5 and 5 kHz. This could be because, even at slow rates and low carrier frequencies, FM is encoded via temporal-envelope cues rather than TFS cues. However, whilst the modelling that disregarded the envelope-phase from all modulation filters predicted slow-rate AM-masking-FM, it suggested this masking should spread up to faster FM rates than was seen behaviorally. It is postulated here that AM masks FM at slow rates due to post-sensory interference, resulting from perceptual grouping of the AM and FM into a single auditory object (possibly due to their similar rate) or from an inability to attend selectively to one modulator in the presence of another modulator at similar rates. This interpretation of masking effects for slow AM and FM rates in terms of "information masking" would not preclude a role of a TFS code in FM detection. In conclusion, the current psychophysical and modelling study indicates that the modulation filter-bank model (as implemented here) can explain some of the features of AM-masking-FM, but it does not provide a unified account of masking effects between AM and FM. Further work is therefore required to clarify the exact nature of the mechanisms responsible for FM detection (particularly at slow rates).

ACKNOWLEDGMENTS

The authors thank Brian C. J. Moore and Laurent Demany for suggestions on interpretation of the data. The authors were supported by two grants from ANR (HEARFIN and HEART projects). This work was also supported by ANR-11-0001-02 PSL* and ANR-10-LABX-0087. The authors wish to thank the associate editor, V. Richards, and two anonymous reviewers for helpful comments on a preliminary version of this manuscript.

Ardoint, M., Lorenzi, C., Pressnitzer, D., and Gorea, A. (**2008**). "Investigation of perceptual constancy in the temporal-envelope domain," J. Acoust. Soc. Am. **123**(3), 1591–1601.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**(6), 2575–2580.

Binns, C., and Culling, J. F. (**2007**). "The role of fundamental frequency contours in the perception of speech against interfering speech," J. Acoust. Soc. Am. **122**(3), 1765–1776.

British Society of Audiology (**2011**). "Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking," Recommended Procedure (British Society of Audiology, Reading, UK), 32 pp.

Dau, T. (**1996**). *Modeling Auditory Processing of Amplitude Modulation* (Bibliotheks- und Informationssystem der Universität Oldenburg, Oldenburg, Germany).

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

Emmerich, D. S., Ellermeier, W., and Butensky, B. (**1989**). "A reexamination of the frequency discrimination of random-amplitude tones, and a test of Henning's modified energy-detector model," J. Acoust. Soc. Am. **85**(4), 1653–1659.

Ernst, S. M. A., and Moore, B. C. J. (**2010**). "Mechanisms underlying the detection of frequency modulation," J. Acoust. Soc. Am. **128**(6), 3642–3648.

Ernst, S. M. A., and Moore, B. C. J. (**2012**). "The role of time and place cues in the detection of frequency modulation by hearing-impaired listeners," J. Acoust. Soc. Am. **131**(6), 4722–4731.

Ewert, S. D. (**2013**). "AFC—A modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the International Conference on Acoustics AIA-DAGA 2013*, March 18–21, Merano, Italy, pp. 1326–1329.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**(3), 1181–1196.

Ewert, S. D., and Dau, T. (**2004**). "External and internal limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**(1), 478–490.

Ewert, S. D., Paraouty, N., and Lorenzi, C. (**2018**). "A two-path model of auditory modulation detection using temporal fine structure and envelope cues," Eur. J. Neurosci. 1–14.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**(4), 1676–1680.

Hsu, A., Woolley, S. M. N., Fremouw, T. E., and Theunissen, F. E. (**2004**). "Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons," J. Neurosci. **24**(41), 9201–9211.

Johannesen, P. T., Pérez-González, P., Kalluri, S., Blanco, J. L., and Lopez-Poveda, E. A. (**2015**). "Predictors of supra-threshold speech-in-noise intelligibility by hearing-impaired listeners," Proc. Int. Symp. Aduit. Audiol. Res. **5**, 125–136, available at https://proceedings.isaar.eu/index.php/isaarproc/article/view/2015-15.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Lorenzi, C., Berthommier, F., and Demany, L. (**1999**). "Discrimination of amplitude-modulation phase spectrum," J. Acoust. Soc. Am. **105**, 2987–2990.

Lorenzi, C., Soares, C., and Vonner, T. (**2001**). "Second order temporal modulation transfer functions," J. Acoust. Soc. Am. **110**, 1030–1038.

Millman, R. E., Lorenzi, C., Apoux, F., Füllgrabe, C., Green, G. G. R., and Bacon, S. P. (**2002**). "Effect of duration on amplitude-modulation masking (L)," J. Acoust. Soc. Am. **111**(6), 2551–2554.

Moore, B. C. J. (**2014**). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss* (World Scientific, Singapore), Chap. 3, pp. 47–79.
Moore, B. C. J., Füllgrabe, C., and Sek, A. (**2009**). "Estimation of the center frequency of the highest modulation filter," J. Acoust. Soc. Am. **125**(2), 1075–1081.
Moore, B. C. J., Mariathasan, S., and Sek, A., (**2018**). "Effects of age on the discrimination of amplitude and frequency modulation for 2- and 10-Hz rates," Acta Acust. united Acust. **104**(5), 778–782.
Moore, B. C. J., and Sek, A. (**1995**). "Effects of carrier frequency, modulation rate and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation)," J. Acoust. Soc. Am. **97**(4), 2468–2478.
Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**(4), 2320–2331.
Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**(1), 327–335.
Palmer, A. R., and Russell, I. J. (**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), 1–15.
Paraouty, N., Ewert, S., Wallaert, N., and Lorenzi, C. (**2016**). "Interactions between amplitude modulation and frequency modulation processing: Effects of age and hearing loss," J. Acoust. Soc. Am. **140**(1), 121–131.
Paraouty, N., and Lorenzi, C. (**2017**). "Using individual differences to assess modulation-processing mechanisms and age effects," Hear. Res. **344**, 38–49.
Paraouty, N., Stasiak, A., Lorenzi, C., Varnet, L., and Winter, I. M. (**2018**). "Dual coding of frequency modulation in the ventral cochlear nucleus," J. Neurosci. **38**(17), 4123–4137.
Ruggles, D., Bharadwaj, H., and Shinn-Cunningham, B. (**2011**). "Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication," Proc. Natl. Acad. Sci. **108**(37), 15516–15521.
Saberi, K., and Hafter, E. R. (**1995**). "A common neural code for frequency- and amplitude-modulated sounds," Nature **374**(6522), 537–539.
Sek, A., Baer, T., Crinnion, W., Springgay, A., and Moore, B. C. J. (**2015**). "Modulation masking within and across carriers for subjects with normal and impaired hearing," J. Acoust. Soc. Am. **138**(2), 1143–1153.
Sek, A., and Moore, B. C. J. (**1995**). "Frequency discrimination as a function of frequency, measured in several ways," J. Acoust. Soc. Am. **97**(4), 2479–2486.
Sheft, S., Shafiro, V., Lorenzi, C., McMullen, R., and Farrell, C. (**2012**). "Effects of age and hearing loss on the relationship between discrimination of stochastic frequency modulation and speech perception," Ear. Hear. **33**(6), 709–720.
Sheft, S., and Yost, W. A. (**2007a**). "Discrimination of starting phase with sinusoidal envelope modulation," J. Acoust. Soc. Am. **121**(2), EL84–EL89.
Sheft, S., and Yost, W. A. (**2007b**). "Modulation detection interference as informational masking," in *Hearing—From Sensory Processing to Perception* (Springer, Berlin, Germany), Part 6, pp. 303–311.
Singmann, H., Bolker, B., Westfall, J., and Aust, F. (**2017**). afex: Analysis of factorial experiments. R package version 0.18-0. https://CRAN.R-project.org/package=afex
Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech-transmission quality," J. Acoust. Soc. Am. **67**(1), 318–326.
Stickney, G., Nie, K., and Zeng, F.-G. (**2005**). "Contribution of frequency modulation to speech recognition in noise," J. Acoust. Soc. Am. **118**(4), 2412–2420.
Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**(2), 1081–1091.
Strickland, E. A., and Viemeister, N. F. (**1996**). "Cues for discrimination of envelopes," J. Acoust. Soc. Am. **99**(6), 3638–3646.
Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**(4), 2040–2050.
Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (**2017**). "A cross-linguistic study of speech modulation spectra," J. Acoust. Soc. Am. **142**(4), 1976–1989.
Wallaert, N., Moore, B. C. J., Ewert, S. D., and Lorenzi, C. (**2017**). "Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation," J. Acoust. Soc. Am. **141**(2), 971–980.
Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (**2017**). "Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age," J. Assoc. Res. Otolaryngol. **18**, 619–633.
Whiteford, K. L., and Oxenham, A. J. (**2015**). "Using individual differences to test the role of temporal and place cues in coding frequency modulation," J. Acoust. Soc. Am. **138**(5), 3093–3104.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. **102**(7), 2293–2298.
Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs" ("The foundations for determining the information capacity of the auditory system"), Acustica **6**(4), 356–381.