

Double-pass consistency for amplitude- and frequency-modulation detection in normal-hearing listeners

Sarah Attia,^{a)} Andrew King, Léo Varnet, Emmanuel Ponsot, and Christian Lorenzi

Laboratoire des systèmes perceptifs (CNRS 8248), Département d'études cognitives, Ecole normale supérieure, Université Paris Sciences et Lettres, 29 rue d'Ulm, 75005 Paris, France

ABSTRACT:

Amplitude modulation (AM) and frequency modulation (FM) provide crucial auditory information. If FM is encoded as AM, it should be possible to give a unified account of AM and FM perception both in terms of response consistency and performance. These two aspects of behavior were estimated for normal-hearing participants using a constant-stimuli, forced-choice detection task repeated twice with the same stimuli (double pass). Sinusoidal AM or FM with rates of 2 or 20 Hz were applied to a 500-Hz pure-tone carrier and presented at detection threshold. All stimuli were masked by a modulation noise. Percent agreement of responses across passes and percent-correct detection for the two passes were used to estimate consistency and performance, respectively. These data were simulated using a model implementing peripheral processes, a central modulation filterbank, an additive internal noise, and a template-matching device. Different levels of internal noise were required to reproduce AM and FM data, but a single level could account for the 2- and 20-Hz AM data. As for FM, two levels of internal noise were needed to account for detection at slow and fast rates. Finally, the level of internal noise yielding best predictions increased with the level of the modulation-noise masker. Overall, these results suggest that different sources of internal variability are involved for AM and FM detection at low audio frequencies.

© 2021 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0006811>

(Received 8 July 2020; revised 20 September 2021; accepted 5 October 2021; published online 15 November 2021)

[Editor: Joshua G Bernstein]

Pages: 3631–3647

I. INTRODUCTION

The auditory perception of amplitude modulation (AM) and frequency modulation (FM) has received substantial interest over the last decades because of the repeated demonstration of the crucial role played by AM and FM cues in robust speech perception (e.g., Shannon *et al.*, 1995; Zeng *et al.*, 2005) and in environmental sound perception (Singh and Theunissen, 2003; Thoret *et al.*, 2020). It is generally assumed that auditory processing of AM and FM overlaps because of “FM-to-AM conversion” in the cochlea (the frequency-dependent attenuation of the FM caused by the tuned cochlear filters resulting in AM cues; Zwicker, 1952; Saberi and Hafer, 1995). Although questions remain as to how slow FM is detected by humans at low audio frequencies (Whiteford and Oxenham, 2015; Paraouty *et al.*, 2016; Paraouty and Lorenzi, 2017; Whiteford *et al.*, 2017; Ewert *et al.*, 2018; Wallaert *et al.*, 2018; King *et al.*, 2019; Moore *et al.*, 2019; Parthasarathy *et al.*, 2020; Whiteford *et al.*, 2020), the consensus that has emerged posits that the perception of AM and fast FM is mediated by AM processing through a cascade of processing stages including: bandpass (cochlear) filtering, fast-acting amplitude compression, half-wave rectification, short-term adaptation with fast time constants, demodulation achieved by lowpass filtering followed

by a modulation filterbank, and a decision stage using either a template-matching strategy or time-averaged statistics such as the envelope power (Viemeister, 1979; Strickland and Viemeister, 1996; Dau *et al.*, 1997a; Dau *et al.*, 1997b; Ewert and Dau, 2000; McDermott and Simoncelli, 2011).

This general model (hereafter referred to as the “modulation-filterbank model”; for recent implementations, see Jepsen *et al.*, 2008; Jørgensen *et al.*, 2013; Bib Berger and Ewert, 2016, 2017; Wallaert *et al.*, 2017, 2018; King *et al.*, 2019; Cabrera *et al.*, 2019) postulates that temporal-modulation cues in sounds are transformed into so-called “neural temporal-envelope” cues (i.e., fluctuations in mean firing rate in auditory neurons) and that fine-timing “temporal fine structure” (TFS) cues (i.e., carrier information) are discarded after demodulation achieved by central (post-cochlear) processes. This type of model incorporates an important source of “inefficiency” in temporal-modulation processing: a Gaussian *internal noise* added to the representation of temporal-envelope cues at the output of modulation filters. Internal noise (Green and Swets, 1966) refers to the accumulation of several sources of variability such as the stochastic nature of neuronal firing, the internal state of the observer organism, or fluctuations in attention (Javel and Viemeister, 2000; Faisal *et al.*, 2008; Amitay *et al.*, 2013). Internal noise is initially included in this model of modulation perception to limit the resolution of the observer and thus auditory sensitivity to temporal

^{a)}Electronic mail: sarah.attia@ens.fr

modulation, that otherwise would be perfect. This implies that the decision processes postulated for temporal-modulation detection use a degraded representation of temporal-envelope cues extracted by modulation filters (i.e., some envelope information is lost along the auditory pathway before final decision making). Surprisingly, little effort has been made to characterize internal noise in temporal-modulation detection, with the exception of the work by Ewert and Dau (2004). This study assessed the contribution of internal noise to AM detection by comparing AM detection thresholds for random- versus frozen-noise carriers. Threshold patterns obtained for the frozen-noise carriers were found to be similar to those obtained for the random-noise carriers when broadband noise was used as a carrier. The absence of any variability in the AM stimulus in the case of the frozen-noise carrier had little impact if any on the listener's decisions. This led the authors to conclude that AM detection performance is affected by a *large amount of internal noise* (relative to external variability) in the temporal-envelope processing pathway. This outcome is consistent with previous work conducted for other perceptual dimensions in the auditory and visual modalities, concluding that the "specific choice generated by an observer on a given trial depends, to a large extent, on a loud source of variability that is not under direct experimental control" (e.g., Barlow, 1956; Green, 1964; Burgess and Colborne, 1988; Pelli and Blakemore, 1990; Neri, 2010).

However, it is important to note that internal noise not only limits observer sensitivity: it is also a source of neural variability that determines the *consistency* of the auditory judgments made by the observer in a given perceptual task. In other words, for a given stimulus, a high level of internal noise is expected to yield more variable responses from trial to trial. To quote Green (1964) on this issue, "internal noise is the limiting factor in a trial-by-trial prediction of the subject's responses." This operational definition of internal noise led to the "double-pass" method (Green, 1964), a psychophysical paradigm aiming at inferring the variance of internal noise for a given task from the agreement of individual responses between two successive trials using identical stimuli. The double-pass technique has been used extensively to assess internal noise for a wide range of perceptual tasks and different sensory modalities (e.g., Green, 1964; Spiegel and Green, 1981; Burgess and Colborne, 1988; Lu and Doshier, 2008; Neri, 2010; Hasan *et al.*, 2012; Vilidaite and Baker, 2017). It also proved useful to characterize perceptual learning in the auditory (Jones *et al.*, 2013) and visual modalities (Lu and Doshier, 2004, 2008; Lu *et al.*, 2006).

Here, the double-pass method was used to assess and compare behavioral consistency—and thus the influence of internal noise—for slow and fast AM and FM detection. According to the modulation-filterbank model described above, AM detection should be mediated by (neural) temporal-envelope cues *irrespective of modulation rate*. This implies that both slow (<10 Hz) and fast AM detection should be limited by the same sources of inefficiency. Consistent with this assumption, the modulation-filterbank

model uses a single source (and variance) of internal noise for temporal-envelope processing irrespective of temporal-envelope rate (Dau *et al.*, 1997a; Dau *et al.*, 1997b). Still, some doubts remain as to whether slow and fast AM are processed by the same mechanisms. For instance, Wright and Dai (1998) showed that the detectability of sinusoidal AM at unexpected rates differs for slow and fast modulation rates. Therefore, it is still unknown whether the same source of internal noise constrains AM processing at slow and fast rates.

For FM, it is generally assumed that detection is also mediated by temporal envelope (rate/place) information at fast modulation rates (higher than 5–10 Hz) or at high carrier frequencies (higher than 4–10 kHz) (Moore and Søk, 1994; Paraouty *et al.*, 2018; Whiteford *et al.*, 2020). As indicated above, this overlap between AM and FM processing is a consequence of "FM-to-AM conversion" in the cochlea: frequency-dependent attenuation of time-varying instantaneous frequency in the FM stimulus caused by the tuned cochlear filters results in AM (i.e., temporal envelope cues) that is encoded *via* fluctuations in the mean discharge rate of auditory-nerve fibers. Still, doubts remain as to how slow (<10 Hz) FM is detected at low carrier frequencies (<2–4 kHz). In that case, FM information is potentially conveyed by changes over time in the pattern of neural phase locking of auditory-nerve fibers or neurons in the low brainstem (Moore and Søk, 1996; Paraouty *et al.*, 2018; Wallaert *et al.*, 2018; Moore and Søk, 2019; Moore *et al.*, 2019; Parthasarathy *et al.*, 2020) because the precision of neural phase locking is constant for frequencies up to about 0.6–2 kHz for most mammals (Palmer and Russell, 1986). Some psychophysical studies also pointed out that TFS cues are not used to detect FM with modulation rates above about 10 Hz, arguing that the central mechanism using neural phase-locking information is sluggish (e.g., Moore and Søk, 1994). This implies that at low carrier frequencies, slow FM detection should be limited by sources of inefficiency distinct from those limiting fast FM and slow and fast AM detection. Ewert *et al.* (2018) implemented this assumption in a "two-path" model of AM and FM detection using different sources (and variances) of internal noise for temporal-envelope (rate/place cues) and TFS (phase-locking cues) processing.

The general goal of the present study was to test whether the modulation-filterbank model described above (i.e., a model of modulation processing using a single source of additive internal noise) was able to account for double-pass consistency data in an AM and FM detection task. More specifically, this study aimed to: (i) assess double-pass consistency of auditory judgments in sinusoidal AM and FM detection using a slow (2 Hz) and a fast (20 Hz) modulation rate, a low (500 Hz) carrier frequency and a bandpass (AM or FM) noise masker (weak or strong in magnitude), and (ii) compare the results with the predictions of the modulation-filterbank model using the same tasks and stimuli.

The double-pass method was used to estimate the consistency of auditory judgments in each experimental

condition. This method (Green, 1964; Burgess and Colborne, 1988; Lu and Doshier, 2008) is typically performed using two passes. Stimuli presented in each pass are masked by an external source of noise, as double-pass consistency is determined by the ratio of internal to external noise (Green, 1964): if external variability outweighs internal variability, then observer consistency should be high. Conversely, if internal variability outweighs external variability, then observer consistency should be low. To estimate double-pass consistency for modulation detection, a 2-interval, 2-alternative forced-choice (2I, 2AFC) detection-in-noise task was run twice with identical sequences of external noise in the two successive passes. External noises used for AM and FM detection were generated in the AM and FM domains, respectively (external variability was therefore introduced in the dimension—AM or FM—relevant for the task). Two levels of the variance of external (modulation) noise were used for each type of noise. The modulation-filterbank model simulated each experimental condition. The bandwidth of the simulated cochlear filters was set to 1 equivalent rectangular bandwidth of a normal-hearing auditory filter (ERB_N; Glasberg and Moore, 1990); envelope phase was preserved at the output of modulation filters tuned below 10 Hz, and for each modulation filter centered at and above 10 Hz, only the Hilbert envelope of the output was passed on. For each experimental condition, the model was run for target modulation strengths set to the mean detection threshold across real participants, and the magnitude of the additive internal noise was varied systematically in order to find the value that minimized the model’s prediction errors estimated in terms of performance and consistency. A unified account of AM and FM perception would be achieved if a single level of the variance of internal noise could minimize the model’s prediction errors estimated in terms of performance *and* consistency in all experimental conditions.

II. METHODS

A. Listeners

Fifteen young normal-hearing listeners aged between 18 and 30 years (mean = 23 years; standard deviation, SD = 3 years) participated in the experiments. Only two of them had previous experience in psychoacoustic measurements. They were recruited through the logistic platform RISC (“Relais d’Information sur les Sciences de la Cognition, UMS CNRS 332”) at Ecole normale supérieure (Paris). Each listener had pure-tone audiometric thresholds ≤20 dB hearing level (HL) between 0.25 and 4 kHz (ANSI, 1996) in both ears, and the audiometric thresholds at the tested frequency (0.5 kHz) ranged between −5 and 20 dB HL (mean = 6.2 dB HL; SD = 5.3 dB HL).

All listeners were fully informed about the goal of the study and provided written consent before their participation. This study was approved by the local ethical committee of University Paris Descartes (CERES, No. IRB: 20143200001072).

B. Stimuli

All stimuli were generated digitally at a sampling rate of 48 kHz using the MATLAB environment (release 2013b). They were presented to the right ear of each participant at a sensation level (SL) of 40 dB using an external soundcard (DAC audio USB Audioengine D3, 24-bit resolution) and Sennheiser HD 280 pro headphones with an impedance of 64 ohms (Old Lyme, CT).

The whole series of tests was conducted in an IAC double-wall soundproof booth and all levels were calibrated using a Bruel and Kjaer 2250 sound-level meter. The latter was also used to calibrate the headphone with an artificial ear (4153 6 cc coupler, IEC Standard-60318-1). Calibration was performed between 0.125 and 4 kHz.

All experiments were based on a 2I, 2AFC paradigm. Each trial was divided into two successive observation intervals containing a 500-Hz pure tone modulated (in amplitude or in frequency) by a modulation-noise masker. Onset and offset of the modulation-noise masker were simultaneous with each 500-Hz tone. In the target interval, a sinusoidal signal at a fixed rate (2 or 20 Hz) was added to the modulation-noise masker before being applied to the 500-Hz pure-tone carrier. The sinusoidal signal had the same duration as the pure-tone carrier. The target and comparison intervals were presented in random order. The two observation intervals were separated by a 400-ms silent interval. The duration of each stimulus was set to 1-s. A 100-ms raised-cosine ramp was applied to the beginning and the end of each stimulus. All stimuli were normalized in terms of their root mean square (rms) power.

1. AM detection

A modulation noise was generated by filtering a 1-s Gaussian white noise $n(t)$ with zero mean below 80 Hz using a 4th-order lowpass Butterworth filter. The SD of the modulation noise, $\sigma_{ext, AM}$, was set to two values: 0.07 or 0.14

modulation-depth units (m.d.u.). The comparison stimuli were sinusoidal carriers (with a frequency $f_c = 500$ Hz) with instantaneous amplitude modulated using the bandpass noise masker, $n(t)$. The target stimuli were sinusoidal carriers (with a frequency $f_c = 500$ Hz) with instantaneous amplitude modulated by the sum of a sinusoidal signal (of modulation rate f_m and AM depth m) and the bandpass noise masker, $n(t)$.

Equation (1) describes the target stimulus $T_{AM}(t)$:

$$T_{AM}(t) = [1 + m \sin(2\pi f_m t + \varphi_{AM}) + \sigma_{ext, AM} n(t)] \sin(2\pi f_c t + \varphi_c). \quad (1)$$

Equation (2) describes the comparison stimulus $C_{AM}(t)$,

$$C_{AM}(t) = \left[1 + \sigma_{ext, AM} n(t) \right] \sin(2\pi f_c t + \varphi_c), \quad (2)$$

where t is the time expressed in seconds (s), m is the AM depth (ranging between 0 and 100%) of the target signal, f_m

is the modulation rate of the target signal expressed in Hertz (2 or 20 Hz), and φ_{AM} and φ_C are the starting phases of the AM target signal and the carrier, respectively. φ_{AM} and φ_C were randomly and independently chosen (between 0 and 2π radians) for each stimulus in the setup experiment (measure of AM detection thresholds) and in the first pass of the main experiment (measure of double-pass consistency for AM detection).

2. FM detection

As for the AM task, a modulation noise was generated by filtering a 1-s Gaussian white noise $n(t)$ with zero mean below 80 Hz using a 4th-order lowpass Butterworth filter. The SD of the modulation noise, σ_{FM}^{ext} , was set to two values:

0.07 or 0.14 modulation-index units (m.i.u.). The comparison stimuli were sinusoidal carriers (with a frequency $f_c = 500$ Hz) with instantaneous frequency modulated using the bandpass noise masker, $n(t)$. The target stimuli were sinusoidal carriers (with a frequency $f_c = 500$ Hz) with instantaneous frequency modulated by the sum of a sinusoidal signal (of rate f_m and magnitude $\beta = \Delta f / f_m$) and the bandpass noise masker, $n(t)$.

Equation (3) describes the target stimulus $T_{FM}(t)$,

$$T_{FM}(t) = \sin[2\pi f_c t + \varphi_c + \beta \sin(2\pi f_m t + \varphi_{FM}) + \sigma_{FM}^{ext} n(t)]. \quad (3)$$

Equation (4) describes the comparison stimulus $C_{FM}(t)$,

$$C_{FM}(t) = \sin\left[2\pi f_c t + \varphi_c + \sigma_{FM}^{ext} n(t)\right], \quad (4)$$

where t is the time expressed in seconds (s), $\beta = \Delta f / f_m$, where Δf is the frequency excursion of the target signal expressed in Hertz, f_m is the modulation rate of the target signal expressed in Hertz (2 or 20 Hz), and φ_{FM} and φ_C are the starting phases of the FM target signal and the carrier, respectively. φ_{FM} and φ_C were randomly and independently chosen (between 0 and 2π radians) for each stimulus in the septup experiment (measure of FM detection thresholds) and in the first pass of the main experiment (measure of double-pass consistency for FM detection).

3. Modulation-noise maskers

For both AM- and FM-detection tasks, the modulation-noise masker, $n(t)$, was refreshed between intervals and trials in the setup experiment (measure of modulation-detection thresholds), and in the first pass of the main experiment (measure of double-pass consistency). The modulation-noise maskers used in the second pass of the main experiment were exactly the same as those used in the first pass of the main experiment (in corresponding trials). Modulation-noise maskers were different across participants (each participant was assigned a unique set of

randomly selected modulation-noise maskers, but the statistics of modulation noises remained globally the same for all participants). Figure 1(A) shows that the long-term modulation spectrum of the modulation-noise maskers is lowpass in shape, cutting around 80 Hz. The effects of the modulation noise on the AM and FM targets are illustrated in the bottom-most panels of Fig. 1, which show the long-term power spectra [Fig. 1(B)] and modulator waveforms [Fig. 1(C)] of the AM and FM target stimuli used in the experiments. Power spectra and modulator waveforms are shown for a 20-Hz sinusoidal AM [Fig. 1(C), top panels] and FM [Fig. 1(B), bottom panels] signal. Figure 1(B) shows that the AM and FM noise maskers introduce audio frequency components on either side of the sidebands of the target signal and that the level of these masking audio frequency components increases when the SD of the modulation noise increases from 0.07 to 0.14 units (m.d.u. or m.i.u.). Figure 1(C) illustrates the corruption caused by the modulation noise on the target sinusoidal modulators in the time domain.

C. Procedure

Two experiments were completed in the same order for each participant: in the first (so-called “setup”) experiment, an adaptive procedure was used to measure individual AM and FM detection thresholds corresponding to three distinct levels of performance. In the second (so-called “main”) experiment, a double-pass consistency task was used to measure the percentage of agreement in the same detection tasks at a fixed level of performance based on individual detection thresholds measured in the first experiment. The whole series of experiments was based on one or two hours per session with several sessions spread out over a 1–4 week period for each participant. Altogether, the experiments lasted on average 14 h per participant. For each task and for each trial, the participant was instructed to choose the interval containing the target sinusoidal modulation.

Each participant was tested for the following 24 experimental conditions: two types of modulation [AM vs FM]; two modulation rates [2 vs 20 Hz]; two levels of the SD of the (external) modulation-noise masker [0.07 vs 0.14 units (m.d.u. or m.i.u.)]; three levels of (targeted) performance [64% ($d' = 0.5$), 76% ($d' = 1$), and 84% ($d' = 1.5$) correct detection]. The two levels of the modulation-noise masker SD were determined in a pilot experiment to yield distinct levels of detectability of modulation targets. These values were also selected to prevent over-modulation. The individual AM and FM detection thresholds were then used to set the modulation strength (i.e., the modulation depth or frequency excursion, respectively) of the target signals in the subsequent double-pass experiments. Hence, the AM and FM detection tasks in the double-pass consistency paradigm were performed at the same level of difficulty.

The double-pass consistency task was conducted with the purpose of estimating PA in the participant responses

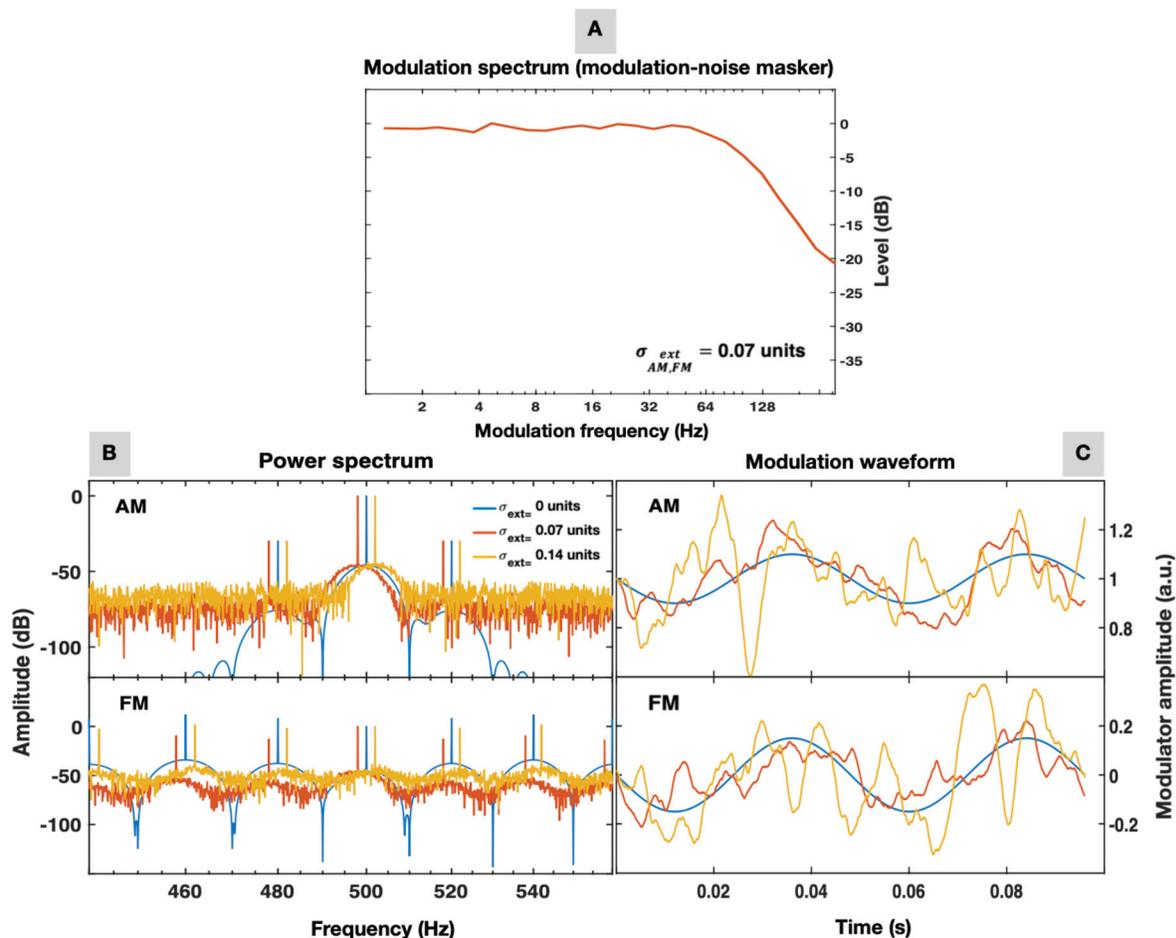


FIG. 1. Stimuli, Modulation-noise maskers and AM/FM targets. (A) Long-term modulation spectrum of AM and FM modulation-noise maskers. The SD of the masker, $\sigma_{ext_{AM,FM}}$, was set to 0.07 units. Modulation spectra were averaged over 100 stimuli. Bottom panels: (B) long-term power spectra of a 20-Hz sinusoidal AM (top panel) and FM (bottom panel) target with (orange and yellow lines) and without (blue lines) AM and FM modulation-noise maskers. Power spectra were offset (± 2 Hz) for clarity. Blue lines (no masker) are shown here to illustrate the detrimental effects caused by modulation-noise maskers; AM and FM stimuli were always presented against a modulation-noise maskers in the setup and main experiments. Here, $\sigma_{ext_{AM,FM}}$, was set to three values: 0 (no masker), 0.07 or 0.14 units (m.d.u. or m.i.u.). For AM target stimuli, modulation depth was set to 10%. For FM target stimuli, frequency excursion was set to 3 Hz. Power spectra were averaged over 10 stimuli. (C) Modulator waveforms of the 20-Hz sinusoidal AM (top panel) and FM (bottom panel) target. Modulator waveforms were obtained for independent samples of modulation noise. *a.u.* stands for arbitrary units.

for corresponding trials in two successive passes. PC was also calculated across the two successive passes for each participant.

1. Setup experiment: Detection thresholds

AM detection thresholds (AM depth, m) and FM detection thresholds (frequency excursion, Δf) were first measured for each participant with adaptive staircases that attempted to converge on 64% ($d' = 0.5$), 76% ($d' = 1$), and 84% ($d' = 1.5$) correct detection using a 2I, 2AFC paradigm, and a weighted one-up, one-down method (García-Pérez, 1998).

Each block consisted of a variable number of trials. Each trial was divided into two successive observation intervals containing the target and comparison stimuli, presented in random order. The bandpass–noise masker was refreshed for each interval and trial. Each observation interval was marked by a red light on the computer screen that was

synchronized with the stimulus presentation. Participants were instructed to report the observation interval containing the target stimulus. Visual feedback was given to the participant after each trial. The dependent variables for these measures were the modulation depth m for the AM detection task and the frequency excursion Δf for the FM detection task. At the beginning of each staircase, the modulation strength of AM and FM targets was set to a supra-threshold level: The AM target was presented at a modulation depth of 80% (-2 dB when expressed as $20 \log m$) and the FM target was presented at a frequency excursion of 8 Hz. At the beginning of each staircase, the tracking variable (the modulation strength of AM and FM targets) changed by a step (a factor) up of four after a wrong response and by a step (a factor) down of two after a correct response. After the first four reversals, the one-down one-up rule was used up to 16 reversals, using different steps to reach 64%, 76%, and 86% of correct responses (García-Pérez, 1998). These 20

reversals were performed for each series of measures and the mean threshold was taken as the geometric mean of the tracking variable during the last 16 reversals. The order of the conditions was randomized across runs.

AM and FM detection thresholds at 2 and 20 Hz were first measured once in the absence of any modulation noise masker in order to familiarize participants with the detection tasks. AM and FM detection thresholds were then measured twice in each condition (i.e., 48 staircases). Two measures of detection threshold were thus carried out in each experimental condition and the final detection threshold for each condition was taken as the mean of these two estimates.

2. Main experiment: Double-pass consistency

The consistency of auditory judgments in AM and FM detection was then measured in each experimental condition and for each participant using the double-pass paradigm developed by Green (1964). Here, AM and FM detection were measured using a 2I, 2AFC paradigm with constant stimuli. In total, each participant was tested over $2 \times 100 = 200$ trials in each of the 24 experimental conditions. In other words, each participant was run twice with identical stimuli, making two successive “passes” of 100 trials in each experimental condition. However, for each trial, the presentation order of the target and comparison stimuli was randomized (and was thus different across the two passes) to avoid contextual effects. The 24 experimental conditions were tested in random order. Thus, 4800 trials were run per participant. These 4800 trials were split into ten (shorter) series of double-pass experiments in which all (24) conditions were completed. Each series was therefore composed of 480 trials: it consisted of a first block of 240 trials in which all (24) conditions were completed ten times in random order, repeated with exactly the same stimuli in a different trial order in a second block of 240 trials (pass 2) after a 5-min break. Short breaks were given to participants between each series of double-pass experiments.

Each block therefore consisted of 240 trials. Five training trials (including random conditions) where modulation targets were presented at a modulation strength yielding a d' of 3 were inserted at the beginning of each block to familiarize the participant with the task and stimuli. The AM depth and frequency excursion used for the training trials were set on an individual basis at twice the minimum threshold value corresponding to a d' of 1.5. For each block, each trial was divided into two successive observation intervals containing the target and comparison stimuli, presented in random order. Each observation interval was marked by a red light on the computer screen that was synchronized with the stimulus' presentation. Participants were instructed to report the observation interval containing the target stimulus. Participants were not given any visual feedback after each trial, except for the first five training trials. Participants' responses to these training trials were not taken into consideration in the assessment of double-pass consistency and percent-correct detection.

For each experimental condition, double-pass consistency was calculated as the percentage of agreement between the participants' responses to corresponding trials within the first and the second pass. The accuracy of responses was also calculated as the percentage of correct responses over the 200 trials [100 trials (pass 1) + 100 trials (pass 2)] corresponding to a given experimental condition.

III. RESULTS

A. Setup experiment (adaptive tasks)

Individual detection thresholds are plotted as a function of the SD of the modulation-noise masker ($\sigma_{AM,FM}^{ext}$) in each panel of Fig. 2. The top [Figs. 2(A) and 2(B)] and bottom panels [Figs. 2(C) and 2(D)] show AM and FM detection thresholds, respectively. Left [Figs. 2(A) and 2(C)] and right panels [Figs. 2(B) and 2(D)] show the 2- and 20-Hz data, respectively. For each panel, detection thresholds are shown for each level of sensitivity targeted by the adaptive procedure (red: $d' = 0.5$; blue: $d' = 1$; green: $d' = 1.5$).

Linear mixed-effect models were run with R (R Core Team, 2018) using the lme4 package (Bates et al., 2012). Independent variables tested were: modulation rate [f_m : 2 or 20 Hz], SD of the modulation noise [$\sigma_{AM,FM}^{ext}$: 0.07 or 0.14

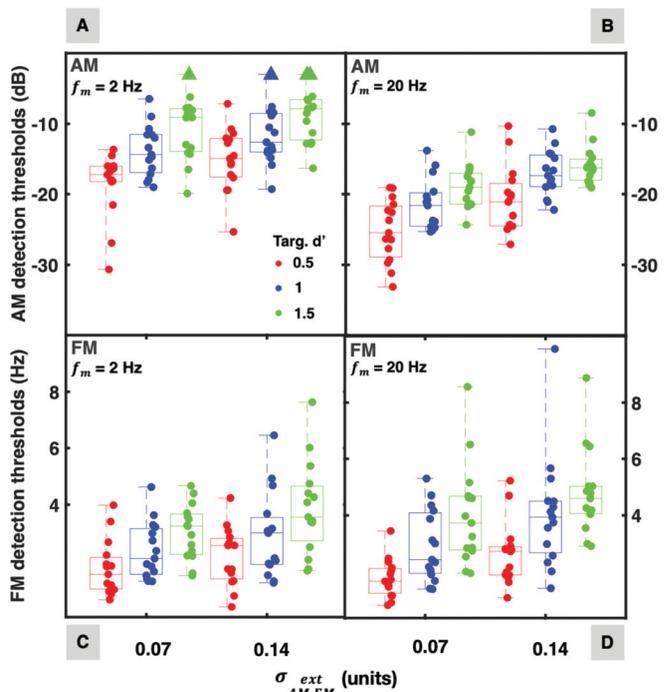


FIG. 2. Individual detection thresholds as a function of the SD of the modulation-noise masker ($\sigma_{AM,FM}^{ext}$). (A)–(D) show AM and FM detection thresholds, respectively. (A), (C), and (C), (D) show the 2-Hz and 20-Hz data, respectively. For each panel, detection thresholds are shown for each level of sensitivity targeted by the adaptive procedure (red: targeted $d' = 0.5$; blue: targeted $d' = 1$; green: targeted $d' = 1.5$). Upward triangles show conditions where participants were not able to perform the AM-detection task. On each box, horizontal lines indicate the median threshold; boxes span the 25th–75th percentiles and vertical lines span the minimum and maximum thresholds.

units (m.d.u. or m.i.u.)), and targeted level of sensitivity [d' : 0.5, 1, or 1.5]. The dependent variable was: (i) AM detection thresholds (expressed in dB, $20 \log m$), or (ii) log-transformed FM detection thresholds. In the statistical model, fixed effects corresponded to f_m , $\sigma_{AM,FM}^{ext}$, and targeted d' , with an interaction term between these variables (to test all significant interactions). Random effects corresponding to individual participants were introduced in the model to take into account individual variability. All categorical variables were expressed using contrast coding. There were no noticeable deviations from homoscedasticity or normality. p -values were obtained by means of type III analysis of variance (ANOVA) on linear-mixed models. The significance level was set at 0.05. The results of these statistical analyses are detailed in Tables I and II of the Appendix.

1. AM detection thresholds

The analysis revealed, as expected: (i) a significant main effect of f_m on AM detection thresholds [$\beta = 6.6$; standard error (SE) = 0.8; $\chi^2(1) = 59.8$; $p < 0.0001$], indicating that AM detection thresholds were significantly lower (better) at 20 Hz: AM detection thresholds were on average 7 dB lower at 20 Hz compared to 2 Hz; (ii) a significant main effect of σ_{AM}^{ext} on AM detection thresholds [$\beta = 4.6$; SE = 0.8; $\chi^2(1) = 4.6$; $p < 0.0001$], indicating that AM detection thresholds increased (worsened) with σ_{AM}^{ext} : AM detection thresholds were on average 7 dB higher at $\sigma_{AM}^{ext} = 0.14$ m.d.u. than at $\sigma_{AM}^{ext} = 0.07$ m.d.u.; (iii) a significant main effect of targeted sensitivity level on AM detection thresholds [$\beta = 6.3$; SE = 0.8; $\chi^2(2) = 53.8$; $p < 0.0001$]. There were no significant interactions between independent variables f_m , σ_{AM}^{ext} , and d' (all $p > 0.05$).

2. FM detection thresholds

The analysis revealed, as expected: (i) a significant main effect of σ_{FM}^{ext} on FM detection thresholds [$\beta = 9.7$; SE = 1.9; $\chi^2(1) = 16.2$; $p < 0.0001$], indicating that FM detection thresholds increased (worsened) with σ_{FM}^{ext} : FM detection thresholds were on average 1.8 times higher at $\sigma_{FM}^{ext} = 0.14$ m.i.u. than at $\sigma_{FM}^{ext} = 0.07$ m.i.u.; (ii) a significant main effect of targeted sensitivity level on FM detection thresholds [$\beta = 8$; SE = 1.9; $\chi^2(2) = 67.8$; $p < 0.0001$]. There was no significant main effect of f_m on FM detection thresholds and no significant interactions between independent variables f_m , σ_{FM}^{ext} , and d' (all $p > 0.05$).

B. Main experiment (double-pass consistency)

During the adaptive tasks, three participants (referred to as A, B, C) reached AM detection thresholds above -5 dB in some experimental conditions, suggesting that they were not able to perform the task. They are shown by upward triangles in Fig. 2. More precisely, one participant (A) could not perform the task in two experimental conditions (2-Hz AM detection, $\sigma_{FM}^{ext} = 0.07$ and 0.14 m.d.u., targeted $d' = 1.5$) and two participants (B, C) could not perform the task in a single experimental condition (participant B: 2-Hz AM detection, $\sigma_{FM}^{ext} = 0.14$ m.d.u., targeted $d' = 1$; participant C: 2-Hz AM detection, $\sigma_{FM}^{ext} = 0.14$ m.d.u., targeted $d' = 1.5$). For these three participants, modulation depth was set to -3 dB in the subsequent (double-pass) constant-stimuli experiment in these corresponding conditions. These participants could perform the task in the double-pass experiment well above chance level (PC $> 75\%$) and PA scores were not affected. All participants could perform the adaptive task in the FM detection experiments.

Figure 3 shows the double-pass consistency and performance data for each participant. The left [Figs. 3(A), 3(C), 3(G), and 3(E)] and right panels [Figs. 3(B), 3(D), 3(H), and 3(F)] show the 2- and 20-Hz data, respectively. The top [Figs. 3(A)–3(D)] and bottom [Figs. 3(E)–3(H)] panels show the data collected using $\sigma_{AM,FM}^{ext}$ of 0.07 and 0.14 units (m.d.u. or m.i.u.), respectively. For each panel, the individual data are shown for each level of sensitivity targeted by the initial adaptive procedure (red: targeted $d' = 0.5$; blue: targeted $d' = 1$; green: targeted $d' = 1.5$). For each experimental condition, a positive correlation was found between PC and PA data: as expected, higher detection performance is associated with higher consistency in listeners' judgments: Bravais-Pearson r ranged between 0.6 and 0.99, except for 20-Hz FM detection, where $r = 0.35$ for $\sigma_{FM}^{ext} = 0.07$ m.i.u. and targeted $d' = 0.5$. Mean PA and PC calculated for each modulation type (AM, FM) and rate (2, 20 Hz) are shown in Table III.

A statistical analysis of double-pass consistency data were conducted using a linear mixed-effect model with R (R Core Team, 2018) using the lme4 package for each dependent variable (PC or PA converted into rationalized arcsine units). Note that this statistical analysis made on empirical PA and PC data were purely descriptive (i.e., it was not testing a specific hypothesis). Modulation type [AM or FM], f_m [2 or 20 Hz], $\sigma_{AM,FM}^{ext}$ [0.07 or 0.14 units (m.d.u. or m.i.u.)] and targeted d' level [0.5, 1 or 1.5] were included as fixed effects together with all interactions between factors; the participants were included as random intercepts. The significance level was set at 0.05 and p -values were obtained using a type III ANOVA on linear-mixed models.

The analysis conducted on PA scores showed that the main effects of modulation type, f_m and $\sigma_{AM,FM}^{ext}$ were not significant (all $p > 0.05$). The analysis only showed a

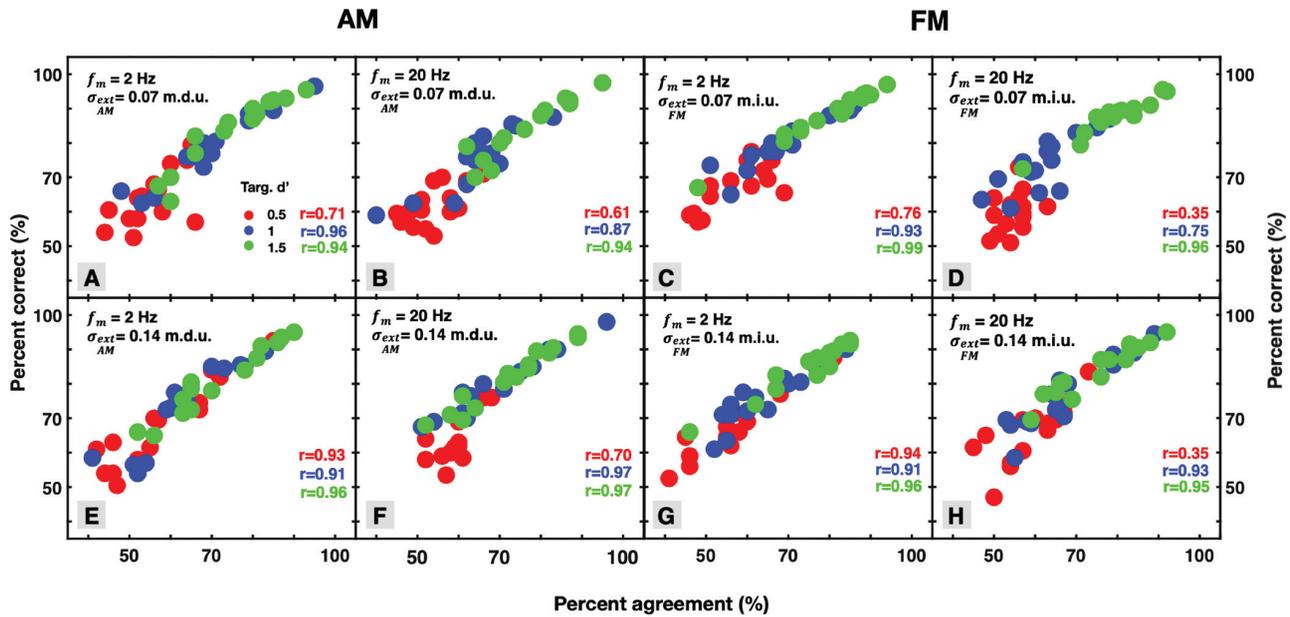


FIG. 3. PC versus PA curves showing double-pass consistency (i.e., percent agreement, PA) and performance (i.e., percent correct, PC) data for each participant. The top and bottom panels show the data collected using σ_{ext} of 0.07 in panels (A), (B), (C), (D) and 0.14 units in panels (E), (F), (G), (H) (m.d.u. or m.i.u.), respectively. For each panel, the individual data are shown for each level of sensitivity targeted by the initial adaptive procedure (red, targeted $d' = 0.5$; blue, targeted $d' = 1$; green, targeted $d' = 1.5$).

significant main effect of targeted d' on PA [$\beta = 10$; $SE = 1.9$; $\chi^2(2) = 55.3$; $p < 0.0001$]. There was also a significant interaction between targeted d' and σ_{ext} [$\beta = -3$; $SE = 1.9$; $\chi^2(1) = 2.7$; $p = 0.046$] indicating that PA was significantly increased when σ_{ext} was highest at targeted d' levels of 0.5 and 1. The analysis did not reveal any significant interactions between independent variables modulation type or f_m (all $p > 0.05$).

The analysis conducted on PC scores confirmed that targeted d' had a significant main effect on PC data [$\beta = 15$; $SE = 2.9$; $\chi^2(2) = 72$; $p < 0.0001$]. The main effects of modulation type, f_m and σ_{ext} , were not significant (all $p > 0.05$). This was expected as listeners were tested at their individual detection threshold targeting a d' of 0.5, 1, and 1.5 in each experimental condition. The analysis did not reveal any significant interaction between independent variables (all $p > 0.05$). The results of these statistical analyses are detailed in Table IV of the Appendix. The outcome of these statistical analyses was not affected by removing the three participants (A, B, and C) that could not perform the task in the setup experiment.

IV. MODEL SIMULATIONS

The general goal of the present study was to test whether the modulation-filterbank model was able to account for the double-pass consistency data collected for young normal-hearing participants in the AM and FM detection tasks described above.

A. Model specifications

The model structure was similar to that used by Wallaert *et al.* (2017, 2018), King *et al.* (2019), and Cabrera *et al.* (2019). This model implemented a cascade of processing stages including peripheral (i.e., cochlear) filtering, half-wave rectification, amplitude compression, short-term adaptation, demodulation (lowpass filtering followed by a modulation filterbank), and a decision stage using a template-matching strategy:

- (1) A bank of five, 1-ERB_N wide gammatone filters, one centered at the carrier frequency f_c of the stimulus, and the remaining four centered at 1 and 2 Cams (units of the ERB_N number scale; Glasberg and Moore, 1990) above and below the f_c of the stimulus.
- (2) A “broken-stick” input-output function for the output of the gammatone filter tuned to the f_c of the stimulus; the function was linear up to a knee-point of 40 dB sound pressure level (SPL) and compressive (using a power law with an exponent of 0.3) above; the remaining four gammatone filters were not submitted to this non-linear transform.
- (3) Half-wave rectification of the output of each of the five gammatone filters.
- (4) High-pass filtering (1st order, 6 dB/oct roll-off, 3-Hz cut-off) of the output of each channel to simulate short-term adaptation (*cf.* Tchorz and Kollmeier, 1999).
- (5) The resulting signal at the output of each channel was passed to a filterbank (Butterworth filters; -6 dB/oct rolloff) with ten logarithmically-spaced channels between 2 and 120 Hz (Moore *et al.*, 2009), each with a

Q factor of 1 (Lorenzi *et al.*, 2001) to decompose the modulations of the processed signals, producing 50 channels; envelope-phase information was preserved at the output of each bandpass modulation filter. For each modulation filter centered below 10 Hz, the waveform at the output of the filter was passed on for further processing, while for each filter centered at and above 10 Hz, only the Hilbert envelope of the output was passed on.

- (6) Independent Gaussian noises were added to the output of each of the 50 channels; this noise had a constant SD σ_{int} expressed in “model” units, m.u. (Wallaert *et al.*, 2017, 2018; King *et al.*, 2019).
- (7) The final decision stage was based on a template-matching strategy (Wallaert *et al.*, 2017). The model generated an “optimal” template at the start of each pass with the modulation strength (the dependent variable: m or Δf) set at a starting value of -2 dB for AM detection and 8 Hz for FM detection, and without any added external and internal noise. The template was calculated as the difference between the internal representations of the target and reference stimuli, channel by channel. On each trial, the target and reference stimulus intervals were correlated (channel by channel) with the template. The interval with the largest correlation coefficient (summed across channels) was selected by the model.

Double-pass consistency was simulated with this general model. The model was initially calibrated with a pure tone presented at 40 dB SPL to map “physical dB SPL” into model units (m.u.) and to set the knee point of the amplitude-compression function (see model description, stage 2). At the beginning of each pass, an internal template was generated in the absence of (external) modulation noise masker and by setting the variance of internal noise to zero. The template was generated using supra-threshold values of the modulation depth for AM and the excursion frequency for FM (-2 dB and 8 Hz, respectively). These supra-threshold values corresponded to those used at the beginning of the adaptive staircases to assess modulation detection thresholds for real participants in the setup experiment. The model was always run with the same stimuli and psycho-physical procedure used for real participants except that (i) the starting phase of target modulations was set to π (instead of being randomized for real participants) because the decision device of the model calculated the correlation between the internal template and the output of modulation filters as in Wallaert *et al.* (2017); (ii) the number of trials that was necessary to obtain reliable estimates of PC and PA scores was increased to 1500 trials per pass (instead of 100 for real participants). Simulations were run using the same Matlab environment (AFC psychoacoustics software, version 1.40.1; Ewert, 2013).

Figure 4 illustrates the model’s behavior for this task, and more specifically the effects of external and internal noise on PC and PA in two experimental conditions: 2-Hz AM detection with σ_{AM}^{ext} set to 0.07 m.d.u. [Fig. 4(A)] and 0.14 m.d.u. [Fig. 4(B)]. In each experimental condition, the

model was simulated for several levels of the internal noise ranging from 0 to 150 m.u. For each level of internal noise, PC and PA were simulated for AM depth varying between -40 dB (1%) and -3 dB (70%) in 1–5 dB discrete steps. Figure 4 shows that PC and PA are strongly related (Lu and Doshier, 2008): a perfect or near-perfect performance is necessarily associated with perfect or near-perfect agreement; in addition, high PC values cannot be associated with low PA values (if $PA < 100\%$, there must be one of the two passes where the observer makes incorrect responses and therefore $PC < 100\%$). Figure 4 also shows that for a given PC level, PA decreases with increasing levels of internal noise: At high levels of internal noise, the PC versus PA curves are slanted; as internal noise decreases, the PC versus PA curves become steeper. Finally, a comparison between the two panels of Fig. 4 shows that PA increases with increasing levels of external noise σ_{AM}^{ext} . In summary, these

preliminary simulations indicate that PA is governed by the ratio between internal-noise and external-noise SD: the higher this ratio, the lower PA.

The model was then run for the 24 experimental conditions described above, and for target modulation strengths (m or Δf) set to the mean detection threshold across participants. For each experimental condition, the model was simulated with a range of values of the SD of the additive internal noise, σ_{int} (between 80 and 300 m.u. for AM detection and between 20 and 160 m.u. for FM detection). For each value of σ_{int} , the absolute error between the model and the participants’ data (the absolute value of the difference in PA and PC between simulated and real data, referred to the “prediction error” below) was calculated.

B. Simulation results

Figure 5 shows the model’s prediction errors (absolute errors between empirical and simulated data) averaged across targeted d' levels as a function of σ_{int} for PA [Figs. 5(A) and 5(C)] and PC [Figs. 5(B) and 5(D)] scores. In each panel, prediction errors are shown for 2-Hz AM (dashed dark-red lines), 20-Hz AM (continuous dark-red lines), 2-Hz FM (dashed black lines), and 20-Hz FM (continuous black lines). Top and bottom panels show the model’s prediction errors for $\sigma_{AM,FM}^{ext} = 0.07$ and 0.14 units (m.d.u. or m.i.u.),

respectively. Figure 5 indicates that in each experimental condition, prediction errors calculated for both PA and PC were substantially affected by changes in σ_{int} . In each experimental condition, the model did not predict PA and PC scores equally well (error curves did not overlap perfectly for PA and PC) but prediction errors dropped below 2–5 points of percentage for both PA and PC for specific values of σ_{int} . The level of σ_{int} that yielded the lowest prediction errors changed with modulation type, modulation rate, and level of the modulation-noise masker. Simulations were run ten times using these values of σ_{int} to check that the model predictions did not change substantially across simulations. The coefficient of variation (CV) was calculated for PA and PC

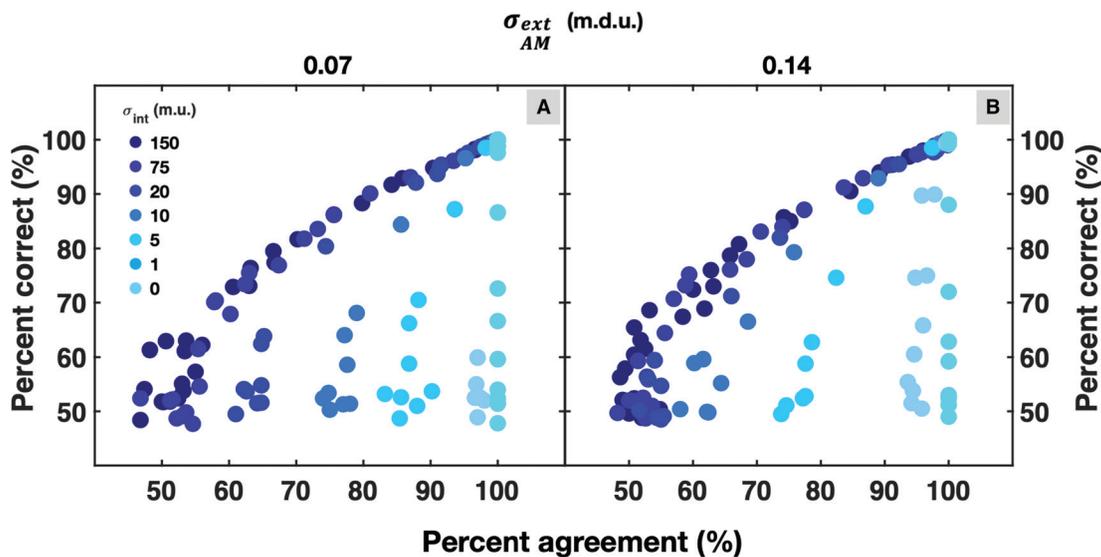


FIG. 4. Predictions of the modulation-filterbank model for the “double-pass” AM detection task. A 2-Hz sinusoidal AM was used as the target to illustrate the model’s behavior for this task. The SD of the modulation-noise masker, $\sigma_{AM,FM}^{ext}$, was set to 0.07 m.d.u. (A) and 0.14 m.d.u. (B). The model was simulated for several levels (i.e., SD, expressed in model units, m.u.) of the additive internal noise. Each dot was calculated from 200 trials. For each level of internal noise (as indicated by the hue code, dark blue corresponding to the highest level), double-pass consistency (percent agreement, PA) and performance (percent-correct detection, PC) were simulated for AM depth varying between -40 dB (1%) and -3 dB (70%) in 1–5 dB discrete steps. For a given PC level, PA decreases with increasing levels of internal noise and it increases with increasing levels of external modulation noise.

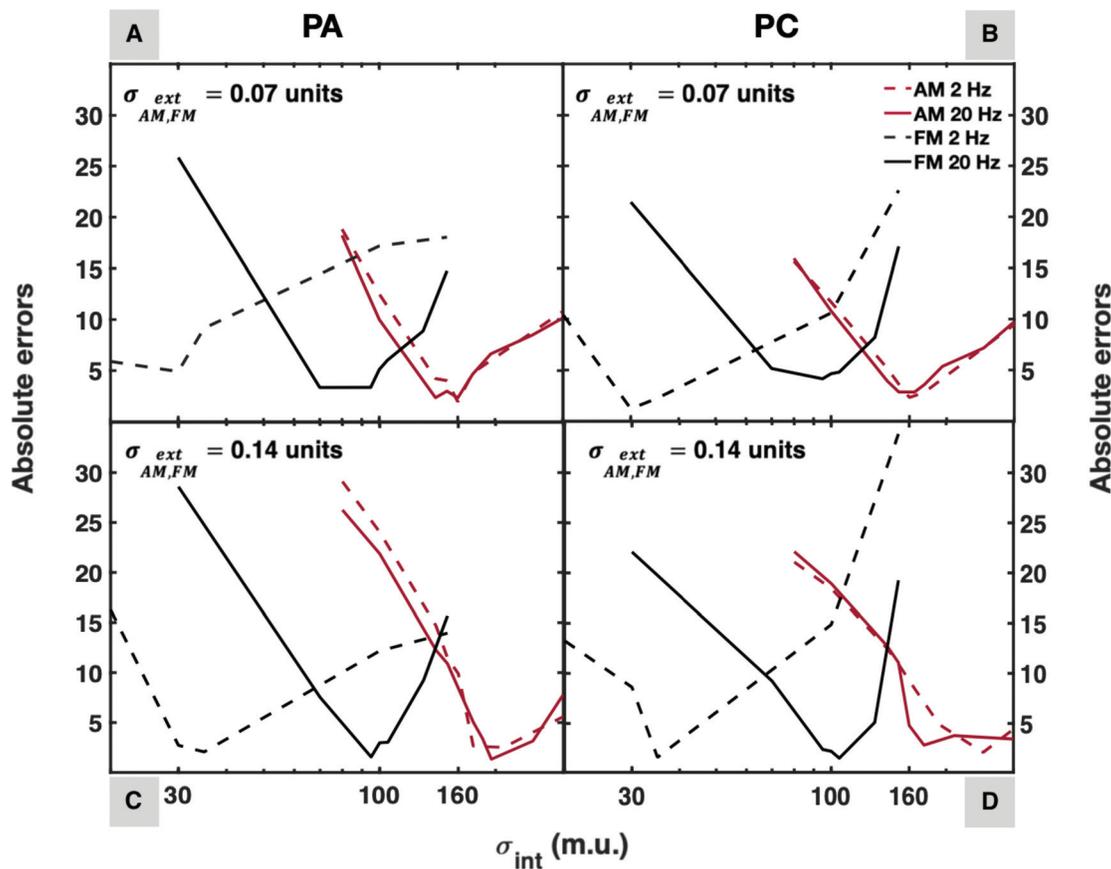


FIG. 5. Prediction errors (average across targeted d' levels) of the modulation-filterbank model as a function of σ_{int} for PA (A and C) and PC (B and D) scores. In each panel, prediction errors are shown for 2-Hz AM (dark red, dashed lines), 20-Hz AM (dark red, continuous lines), 2-Hz FM (black, dashed lines), and 20-Hz FM (black, continuous lines). Top (A and C) and bottom (B and D) panels show the model’s prediction errors for $\sigma_{AM,FM}^{ext} = 0.07$ and 0.14 units (m.d.u. or m.i.u.), respectively. The x axis is logarithmic.

in each experimental condition. CV was generally less than 0.5, indicating that model predictions were quite reproducible. Overall, the level of σ_{int} that yielded the lowest prediction errors for PA and PC was lower for FM than AM, and lower for 2-Hz FM than for 20-Hz FM; for each type of modulation, it was generally lower at the lowest level of the modulation-noise masker. For each experimental condition, values of σ_{int} lower than this level were associated with simulated PA and PC scores higher than empirical scores. Conversely, values of σ_{int} higher than this level were associated with simulated PA and PC scores lower than empirical scores.

The following lines describe the effects of the modulation-noise masker on simulated PA and PC scores. When $\sigma_{ext_{AM}}$ was lowest (0.07 m.d.u.), prediction errors were lower than 5% points for both PA and PC for 2- and 20-Hz AM when $\sigma_{int} = 160$ m.u. When $\sigma_{ext_{AM}}$ was increased to 0.14 m.d.u., higher levels of internal noise were required to reach a minimum in prediction errors for both PA and PC for 2- and 20-Hz AM ($\sigma_{int} = 175$ and 250 m.u., respectively). When $\sigma_{ext_{FM}}$ was lowest (0.07 m.i.u.), a level of internal noise of 30 m.u. was associated with prediction errors lower than 5 points of percentage for both PA and PC for 2-Hz FM; in comparison, a level of 100 m.u. was required to reach a minimum in prediction errors for both PA and PC for 20-Hz FM. When $\sigma_{ext_{FM}}$ was increased to 0.14 m.i.u., a higher level of internal noise was required for 2-Hz FM only ($\sigma_{int} = 35$ m.u.). It is noteworthy that an increase in

$\sigma_{ext_{AM}}$ was associated with an over-estimation of PA scores for AM; an increase in $\sigma_{ext_{FM}}$ was associated with an over-estimation of PC for FM. Therefore, the modulation-noise masker had more impact on simulated PA and PC scores than on real scores.

In Fig. 5, model's prediction errors were calculated with a range of values of the SD of the additive internal noise, σ_{int} , and averaged across the three d' levels targeted by the initial adaptive procedure. Figure 6 shows the simulated PA and PC scores for each targeted d' level (red: targeted $d' = 0.5$; blue: targeted $d' = 1$; green: targeted $d' = 1.5$) when a fixed level of internal noise was used, that is for $\sigma_{int} = 160$ m.u. In each panel, the simulated data (open circles) are plotted along with the empirical data (mean across participants; closed circles). As indicated above, the model predicted accurately PA and PC scores for 2- and 20-Hz AM when $\sigma_{ext_{AM}}$ was lowest (0.07 m.d.u.). For 2-Hz AM, increasing $\sigma_{ext_{AM}}$ to 0.14 m.d.u. yielded accurate predictions for PC scores, but PA scores were over-estimated (i.e., better than observed in real participants). For 20-Hz AM, both PA and PC scores were over-estimated when $\sigma_{ext_{AM}} = 0.14$ m.d.u. The model failed to account for FM data. For 2-Hz FM, the model predicted PA and PC scores at or just above chance level (<60%). For 20-Hz FM, the predicted PC scores increased up to about 70% for a targeted d' of 1.5; however, predicted PA scores remained close to

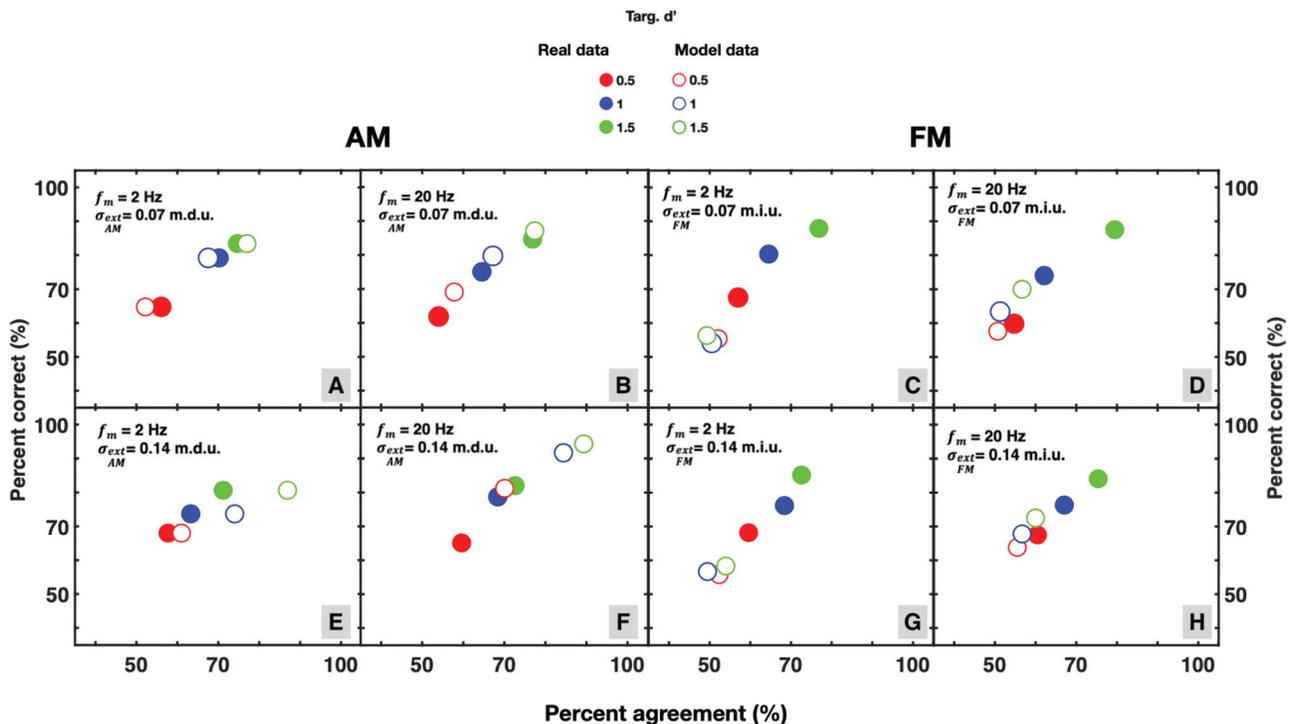


FIG. 6. PC versus PA data for AM and FM detection. Each panel shows the simulation data obtained with $\sigma_{int} = 160$ m.u. plotted along with the empirical data (mean across participants; closed circles). In each panel, the simulation (closed circles) and real data (open circles) are shown for each level of sensitivity targeted by the initial adaptive procedure (red, targeted $d' = 0.5$; blue, targeted $d' = 1$; green, targeted $d' = 1.5$).

chance level, inconsistent with the real data. Thus, the model failed when the internal noise was fixed. In other words, different levels of internal noise were required to fit all the data.

Taken together, these results indicate that the modulation-filterbank model was able to reproduce accurately double-pass consistency data for both slow and fast AM at the lowest level of the modulation-noise masker (here, $\sigma_{AM}^{ext} = 0.07$ m.d.u.) but

it was more sensitive to changes in the level of the modulation-noise masker than the real participants. Still, the modulation-filterbank model could not account for double-pass consistency data for both AM and FM detection using a single source (and level) of additive internal noise. In addition, two different levels of internal noise were needed to account for the slow and fast FM data.

V. DISCUSSION

A. Separate source(s) of internal variability for AM and FM detection at low carrier frequencies

1. Separate sources of internal noise for AM and FM detection

Our crude implementation of the “modulation-filterbank model” using a single source (and thus, a single level) of additive internal noise gave a good account of consistency and detection performance of listeners’ judgments (mean PA and PC data across listeners) for both slow (2 Hz) and fast (20 Hz) AM when external variability conveyed by the modulation-noise masker was relatively low ($\sigma_{AM}^{ext} = 0.07$ m.d.u.). The model could also account for slow and fast FM detection but for levels of the internal noise smaller than that yielding best model predictions for AM. Moreover, different levels of internal noise were required to predict performance and consistency of listeners’ judgments for slow and fast FM detection.

Taken together, these results are consistent with the idea that the same sources of sensory and cognitive variability constrain slow and fast AM processing. These results add to those of recent studies demonstrating the capacity of the modulation-filterbank model to account for a large set of detection, discrimination or identification data collected in psychophysical tasks where listeners are expected to rely on temporal-envelope cues (e.g., Biberger and Ewert, 2016). It is important to note that the present study tested for the first time the ability of the modulation-filterbank model to reproduce the consistency of human auditory judgments in AM detection tasks in addition to accuracy.

Our model could not provide a unified account of AM and FM detection at a low audio frequency (500 Hz). These results are therefore inconsistent with the notion that slow and fast FM are detected via a simple frequency-to-place (FM-to-AM) conversion mechanism operating at the cochlear level (Whiteford et al., 2017, 2020). More precisely, these results show that FM-to-AM conversion cues resulting from cochlear filtering cannot be used by a modulation filterbank to detect either slow or fast FM. Other mechanisms and thus, other

sources of internal variability make a noticeable contribution to FM detection irrespective of FM rate. Recent electrophysiological data collected on guinea pigs revealed that both slow and fast FM are well encoded in the phase-locked discharge patterns of cochlear-nucleus neurons (Parouty et al., 2018). Therefore, these sources of variability may be related to those constraining the coding of TFS cues via neural phase locking in auditory-nerve fibers and brainstem neurons (e.g., Moore and Søk, 1992). Alternatively, these sources of variability may be related to those constraining a more central mechanism distinct from those involved in monaural AM detection. It is still possible that the latter may be using out-of-phase temporal-envelope cues at different cochlear place, but they would require a neural machinery different from that used to detect AM (this issue is discussed further below). The effects of the mechanism responsible for FM detection on behavioral variability—whether it corresponds to neural phase locking to TFS cues or not—may be estimated within the present framework in terms of an “equivalent” internal noise in the temporal-envelope domain. The level of this “equivalent” internal noise is smaller (by a factor 1.5–5) than that constraining AM detection. This suggests that this mechanism is more efficient than that operating on temporal-envelope fluctuations at the output of cochlear filters.

2. Mechanisms of FM detection at low carrier frequencies

It is perhaps surprising that the modulation-filterbank model proved unable to predict FM detection scores (PC) above chance level in the present experimental conditions, knowing that FM detection thresholds at fast (20 Hz) rate could be predicted relatively well in previous work (Wallaert et al., 2018; King et al., 2019) using a comparable implementation of the modulation-filterbank model. The reason for this might be the addition of the (FM) modulation-noise masker in the present FM detection experiments, a crucial aspect of the double-pass design (Green, 1964). Previous work used either no masker or a deterministic (sinusoidal) AM masker when measuring and predicting FM detection thresholds. It follows that the stochastic fluctuations introduced by the FM-noise masker prevented the template-matching decision module of the current model to make efficient use of the temporal-envelope cues resulting from FM-to-AM conversion at the output of cochlear filters.

Additional model simulations were carried out to test whether the model succeeds in accounting for modulation-detection data in the absence of modulation-noise masker, but fails when a modulation-noise masker is present. The current implementation of the modulation-filterbank model was simulated with or without external modulation noise in the adaptive tasks (cf. setup experiment). Here, we used the level of variance of internal noise $\sigma_{int} = 160$ m.u. that minimized the model’s prediction errors for the double-pass experiment for 2- and 20-Hz AM in the presence of low external modulation noise ($\sigma_{AM}^{ext} = 0.07$ m.d.u.; cf. Fig. 5).

When there was no external modulation noise, the model

accounted for previously published data for slow and fast AM and fast FM but failed for slow FM. When external modulation noise was applied, the model accounted successfully for the present AM data, but failed to account for the present FM data.

More precisely, our model reproduced unmasked AM-detection thresholds at 2 and 20 Hz very well (−18.2 and −25.8 dB, respectively, when targeted $d' = 1$; for comparison: −17 and −24 dB for real listeners in King *et al.* (2019) in comparable conditions) and unmasked FM-detection thresholds at 20 Hz (3.3 Hz when targeted $d' = 1$; for comparison: 2.3 Hz for real listeners in King *et al.* (2019) in comparable conditions). As in King *et al.* (2019), our model underestimated real unmasked 2-Hz FM-detection thresholds by a large amount (7.4 Hz when targeted $d' = 1$; for comparison: 1.8 Hz for real listeners in King *et al.*, 2019, in comparable conditions).

Our model was also run using 2 and 20-Hz AM and FM masked by low external modulation noise ($\sigma_{ext}^{AM,FM} = 0.07$

units) using $\sigma_{int} = 160$ m.u. Our model reproduced the present masked AM-detection thresholds at 2 and 20 Hz very well (−14.2 and −20.7 dB, respectively, when targeted $d' = 1$; for comparison: −13.8 and −21.4 dB for the real listeners tested here). Not surprisingly, our model underestimated the present masked FM-detection thresholds by a substantial amount at 2 and 20 Hz (12 and 6 Hz, respectively, when targeted $d' = 1$; for comparison: 2.4 and 3 Hz, respectively, for the real listeners tested here). These simulation data are shown in Table V (simulation data are also shown for targeted d' of 0.5 and 1.5).

In summary, the modulation-filterbank model as implemented here could successfully account for slow and fast-AM detection measured in the presence or absence of modulation maskers with a random structure. The modulation-filterbank model could also account for fast-FM detection in the absence of masker or in the presence of a deterministic masker (e.g., a sinusoidal AM) on the sole basis of FM-to-AM conversion cues resulting from cochlear filtering. In other words, the model could use the out-of-phase envelope patterns elicited by FM at the output of cochlear filters tuned to frequencies below and above the carrier frequency to detect fast FM at a performance level similar to that of real listeners. The modulation-filterbank model therefore gives a parsimonious account of the fast-FM data measured in these conditions: no additional mechanism—such as a cross-channel mechanism—is required to account for these behavioral data. However, the model clearly fails when fast FM is presented against a masker with random structure, indicating that FM-to-AM conversion does not convey robust cues for FM detection in noisy conditions and that an additional mechanism distinct from that involved in AM detection is required to account for performance and sensitivity of real listeners in this case.

It has been assumed that even at low carrier frequencies (<2–4 kHz) where phase locking information is available, FM detection at high rates is mediated by temporal-

envelope cues resulting from FM-to-AM conversion at the output of cochlear filters because auditory processing of TFS cues may be too sluggish (e.g., Moore and Sèk, 1994). Still, electrophysiological (i.e., single-unit) data collected on guinea pigs indicate that fast FM carried by low-frequency tones is accurately encoded in terms of neural phase-locking to TFS cues in the low brainstem, and more precisely, in the ventral cochlear nucleus (Paraouy *et al.*, 2018). Thus, it might be the case that at threshold, FM detection at high rates is mainly mediated by (salient) temporal-envelope cues resulting from FM-to-AM conversion when FM signals are presented in quiet or masked by deterministic (that is, non-random) sounds such as a sinusoidal AM masker, as in Wallaert *et al.* (2018) or King *et al.* (2019); however, neural TFS cues—although weaker because of sluggish processing at more central stages—may come into play (in addition to temporal-envelope cues) when fast-FM signals are corrupted by concurrent sounds showing random fluctuations in instantaneous amplitude and frequency, as in the present study. In other words, temporal-envelope cues resulting from FM-to-AM conversion may be useful to detect fast FM but they may not be very useful in the presence of masking sounds with random structure. This interpretation is compatible with the idea that processing of TFS cues at a relatively central stage of the auditory system is sluggish: it mainly suggests that neural TFS cues evoked by fast FM—although weaker because of such central limitations—may help listeners when temporal-envelope cues cannot be used efficiently by decision-making mechanisms.

Still, it is not possible to exclude the possibility that a central “cross-channel” mechanism using out-of-phase envelope cues at different cochlear place is responsible for both slow and fast FM detection. Neurophysiological and modelling work is required to determine the plausibility of such a mechanism.

B. Characterization of internal noise in the temporal-envelope domain

1. Gaussian internal noise?

Internal noise was modelled as additive and it followed a Gaussian distribution. Still, it may be the case that other statistical distributions could yield better predictions. For instance, Neri (2013) showed that internal noise for visual detection tasks followed a leptokurtic distribution (a distribution with kurtosis larger than that of a normal distribution) instead of a normal one. Such a distribution might result from slow changes in the variance of the distribution over time (see Goris *et al.*, 2014, for the neural basis of this phenomenon). Further work is thus warranted to explore this possibility.

2. Nature of internal noise

In the model, internal noise was added at the output of modulation filters, before the decision device. From the current experiments, it is unfortunately impossible to infer whether this source of internal variability is of sensory or

cognitive (i.e., attentional, memory) origin, and whether it results from a single source of neural variability (e.g., spontaneous activity in central neurons tuned to specific temporal-envelope rates) or from many separate sources located at different (peripheral and central) stages of the auditory system.

The double-pass procedure was developed to assess the influence of the total amount of internal noise, both additive and multiplicative, relative to external noise, on perceptual processes (Green, 1964; Spiegel and Green, 1981; Lu and Doshier, 2008). The model used an additive (that is a fixed-variance) internal noise only. It may be that the relative inability of the model to account for the effects of external noise on PA and PC scores using a single level of additive internal noise resulted from the absence of multiplicative internal noise in the model, i.e., an additional source of internal variability whose variance would be proportional to the envelope power of the stimuli. This would be consistent with the fact that the SD of internal noise that yielded best model predictions increased with the SD of the modulation-noise masker. Indeed, Ewert and Dau (2004) previously showed that a multiplicative noise constrains temporal-envelope processing and is required to simulate the Weber's law behavior for AM-depth discrimination and the absence of certain carrier effects (i.e., random versus frozen broadband noise carriers) on AM detection. Additional work is thus necessary to characterize further the effects of external variability on AM detection and the contribution of a multiplicative internal-noise source to consistency in AM detection.

C. Limitations and novelty of the present approach

As pointed out in Sec. III (see Fig. 4), response consistency and performance co-vary (Lu and Doshier, 2008). Therefore, the same conclusion (i.e., separate sources of internal noise, and thus separate mechanisms constrain AM and FM detection) would have been reached by measuring performance only, as indicated by comparable (although not identical) performance and consistency curves in Fig. 5. This could be interpreted as an important limitation of the present approach based on the double-pass paradigm. Still, it is crucial to keep in mind that the primary goal of this study was to give a *unified* account of AM and FM perception *both* in terms of response consistency and performance. This is why consistency and performance curves shown in Fig. 5 should not be read separately but in conjunction. The important result here is that, for a given level of external noise, the same level of internal noise (e.g., 160 m.u.) yields the best predictions for *both* consistency and performance data. It was not clear from the outset that the modulation-filterbank model could account for both performance and consistency of human judgments in slow and fast AM detection tasks, as pointed out by the simulation results shown in Fig. 4. Demonstrating this was an important step confirming further the validity of this modelling approach for AM processing, adding to the outcome of previous studies

demonstrating the capacity of the modulation-filterbank model to account for a large set of psychophysical data collected in tasks where listeners are expected to rely on temporal-envelope cues (e.g., Biberger and Ewert, 2016). This demonstration was also a prerequisite when interpreting the incapacity of the modulation-filterbank model to account for FM detection (consistency and performance) data for slow and fast FM presented against noisy maskers. In particular, it was not anticipated that the modulation-filterbank model would not be able to account for consistency and performance data in the fast FM condition because previous simulation studies of Wallaert *et al.* (2018) and King *et al.* (2019) showed that this model could predict successfully fast FM detection thresholds in the presence or absence of a sinusoidal AM masker.

D. Implications for the assessment of the perceptual consequences of ageing and cochlear damage

The present study was conducted with young normal-hearing participants. Some studies aiming to model AM detection and discrimination for elderly listeners and listeners with sensorineural hearing loss suggest that the variance of the additive internal noise may increase substantially (up to a factor 6–10) because of aging and cochlear lesions (Derleth *et al.*, 2001; Ives *et al.*, 2014; Paraouty *et al.*, 2016; Wallaert *et al.*, 2017, 2018). From a behavioral perspective, this increase in internal noise should yield a decrease in the consistency of the responses of elderly and hearing-impaired persons, as measured in the current double-pass paradigm. However, to the best of our knowledge, this consistency has never been investigated in these two populations. The degree of consistency of responses and the estimation of internal noise could be of clinical interest by complementing the performance and sensitivity measures traditionally used to characterize suprathreshold deficits caused by sensorineural hearing loss and presbycusis. These measures may be related to the degree of neural deafferentation (or “synaptopathy”) caused by aging and noise exposure (Lopez-Poveda, 2014; Marmel *et al.*, 2015), which is supposed to affect mainly the processing of high sound amplitudes, and, consequently, the neural coding of temporal-envelope cues (Furman *et al.*, 2013; Bharadwaj *et al.*, 2014). Consistent with this hypothesis, a recent modelling study by Goodman *et al.* (2018) indicates that massive loss of auditory-nerve fibers results in an increase in neural variability and alterations of AM coding at the low brainstem (cochlear nucleus) level.

In that respect, it is possible that a better understanding of the impact of aging and cochlear lesions on double-pass consistency and internal noise would allow for a more accurate estimation of the limits of signal amplification and speech-enhancement systems applied by current digital hearing aids.

VI. SUMMARY AND CONCLUSIONS

Response consistency and performance in a modulation-detection task were estimated for young,

normal-hearing participants using a double-pass paradigm and sinusoidal AM and FM targets masked by a modulation-noise masker. The double-pass data were simulated using a computational model of temporal-envelope processing implementing a peripheral filterbank, half-wave rectification, instantaneous amplitude compression, adaptation, a filterbank of bandpass modulation filters, an additive internal noise, and a template-matching decision device.

The results showed that:

- (1) Distinct levels (i.e., SD) of additive internal noise were required to reproduce AM and FM data but a single level of internal noise could account for the 2- and 20-Hz AM data.
- (2) The level of internal noise that yielded the best model predictions increased with the SD of the modulation-noise masker.
- (3) The level of internal noise yielding best model predictions for FM detection was smaller (by a factor 1.5–5) than the one yielding best model predictions for AM detection. In addition, two different levels of internal noise were needed to account for slow and fast FM detection.

These results suggest that distinct sources of internal variability are involved for AM and FM detection at low audio frequencies. They also suggest that AM detection may be constrained by an additional source of internal noise whose variance is proportional to the envelope power of the stimuli and that the mechanism involved in FM detection is more efficient than the one operating on temporal-envelope fluctuations at the output of cochlear filters.

In conclusion, the current study based on the double-pass paradigm showed that the modulation filterbank-model is able to give a unified account of slow and fast AM detection both in terms of response consistency and accuracy. This demonstrates further the capacity of this modelling approach to account for a large set of perceptual data. However, an additional mechanism distinct from that involved in AM detection must be considered to account for slow and fast FM detection at low carrier frequencies.

ACKNOWLEDGMENTS

This work was supported by two grants from the Collège National des Audioprothésistes (France) and Entendre SAS. This work was also supported by ANR-17-EURE-0017. The authors wish to thank warmly Peter Neri for helpful discussions about the double-pass paradigm, J. Bernstein, and two anonymous reviewers for very constructive feedback on the present article. This work was supported by two grants from the College National des Audioprothésistes (France) and Entendre SAS. This work was also supported by ANR-17-1185 EURE-0017, ANR-16-CE28-0016 and ANR-19-CE28-0010-01.

APPENDIX

See Tables I–V show the results of the statistical analyses.

TABLE I. Analysis of deviance (type III Wald chi-square tests) showing p -values from lmer models (lme4 package) including AM-detection thresholds as dependent variable, and modulation rate f_m , targeted sensitivity level (d'), $\sigma_{AM,FM}^{ext}$ as independent variables. df, degree of freedom.

AM-detection thresholds	χ^2	df	p
Intercept	656.28	1	2.20×10^{-16}
σ_{ext}	28.4702	1	9.51×10^{-8}
f_m	59.8814	1	1.00×10^{-14}
d'	53.8709	2	2.00×10^{-12}
$\sigma_{AM}^{ext} * f_m$	0.8539	1	0.35
$\sigma_{AM}^{ext} * d'$	2.0002	2	0.37
$f_m * d'$	1.9334	2	0.38
$\sigma_{AM}^{ext} * f_m * d'$	0.5336	2	0.77

TABLE II. Analysis of deviance table (type III Wald chi-square tests) showing p -values from lmer models (lme4 package) including FM-detection thresholds as dependent variable, and modulation rate f_m , targeted sensitivity level (d'), $\sigma_{AM,FM}^{ext}$ as independent variables. df stands for degree of freedom.

FM detection thresholds	χ^2	df	P
Intercept	18.8811	1	1.39×10^{-5}
$\sigma_{AM,FM}^{ext}$	16.2447	1	5.67×10^{-5}
f_m	0.5465	1	0.46
d'	67.8274	2	1.86×10^{-15}
$\sigma_{FM}^{ext} * f_m$	1.9553	1	0.16
$\sigma_{FM}^{ext} * d'$	1.4967	2	0.47
$f_m * d'$	1.5699	2	0.45
$\sigma_{FM}^{ext} * f_m * d'$	0.8287	2	0.66

TABLE III. Mean percent-agreement (PA, in %) in each condition. Percent correct (PC, in %) is shown between parentheses. Data averaged across the 15 participants.

Conditions	AM		FM		
	$f_m = 2$ Hz	$f_m = 20$ Hz	$f_m = 2$ Hz	$f_m = 20$ Hz	
$\sigma_{AM,FM}^{ext} = 0.07 u$	$d' = 0.5$	56 (65)	54 (62)	57 (68)	55 (60)
	$d' = 1$	70 (79)	64 (75)	69 (80)	62 (74)
	$d' = 1.5$	75 (83)	77 (85)	80 (88)	79 (88)
$\sigma_{AM,FM}^{ext} = 0.14 u$	$d' = 0.5$	58 (68)	60 (65)	57 (68)	60 (68)
	$d' = 1$	64 (74)	68 (79)	65 (76)	67 (76)
	$d' = 1.5$	71 (81)	73 (82)	75 (85)	75 (84)

Amitay, S., Guiraud, J., Sohoglu, E., Zobay, O., Edmonds, B. A., Zhang, Y. X., and Moore, D. R. (2013). "Human decision making based on variations in internal noise: An EEG study," *PloS one* **8**, e68928.

ANSI (1996). *Determination of Occupational Noise Exposure and Estimation of Noise-Induced Hearing Impairment* (American National Standards Institute, New York).

Barlow, H. B. (1956). "Retinal noise and absolute threshold," *J. Opt. Soc. Am.* **46**, 634–639.

TABLE IV. Analyses of deviance tables (type III Wald chi-square tests) showing p -values from lmer models (lme4 package) including PA and PC (between parentheses) as dependent variables, and M_t (modulation type: AM or FM), modulation rate f_m , targeted sensitivity level (d'), $\sigma_{AM,FM}^{ext}$ as independent variables. df stands for degree of freedom.

PA (PC)	χ^2	df	p
Intercept	424 (596)	1 (1)	2.20×10^{-16} (2.20×10^{-16})
f_m	0.38 (0.99)	1 (1)	0.53 (0.32)
d'	55.28 (71.98)	2 (2)	9.90×10^{-13} (2.33×10^{-16})
M_t	0.044 (0.42)	1 (1)	0.83 (0.52)
$\sigma_{AM,FM}^{ext}$	2.76 (1.17)	1 (1)	0.096 (0.28)
$f_m * d'$	3.77 (2.58)	2 (2)	0.15 (0.27)
$f_m * M_t$	0.011 (1.17)	1 (1)	0.91 (0.27)
$d' * M_t$	1.34 (1.83)	2 (2)	0.51 (0.40)
$f_m * \sigma_{AM,FM}^{ext}$	0.58 (0.0098)	1 (1)	0.44 (0.92)
$d' * \sigma_{AM,FM}^{ext}$	6.13 (4.15)	2 (2)	0.046 (0.12)
$M_t * \sigma_{AM,FM}^{ext}$	0.0070 (1.21)	1 (1)	0.93 (0.27)
$f_m * d' * M_t$	0.14 (0.22)	2 (2)	0.93 (0.89)
$f_m * d' * \sigma_{AM,FM}^{ext}$	4.33 (4.70)	2 (2)	0.11 (0.095)
$f_m * M_t * \sigma_{AM,FM}^{ext}$	0.1015 (1.57)	1 (1)	0.75 (0.21)
$d' * M_t * \sigma_{AM,FM}^{ext}$	0.0052 (1.37)	2 (2)	0.99 (0.50)
$f_m * d' * M_t * \sigma_{AM,FM}^{ext}$	0.2211 (1.74)	2 (2)	0.89 (0.42)

TABLE V. Simulated modulation-detection thresholds (thresholds averaged across 50 simulations per condition) with and without external modulation noise. Empirical (real) data are shown between parentheses. For conditions without external modulation noise, empirical data are taken from King *et al.* (2019). All thresholds correspond to 71% correct scores.

Real vs Model data		AM detection thresholds (dB)		FM detection thresholds (Hz)	
f_m (Hz)		2	20	2	20
$\sigma_{AM,FM}^{ext} = 0$ (s.u.)		-18.2 (-17)	-25.8 (-24)	7.4 (1.8)	3.3 (2.3)
$\sigma_{AM,FM}^{ext} = 0.07$ (s.u.)	$d' = 0.5$	-18.9 (-18.5)	-26.4 (-25.2)	7.3 (1.7)	3.4 (1.7)
	$d' = 1$	-14.2 (-13.8)	-20.7 (-21.4)	12.1 (2.4)	5.9 (3)
	$d' = 1.5$	-10.2 (-10.6)	-17.2 (-18.9)	17.5 (3.1)	8.3 (3.4)

Bates, D., Maechler, M., and Bolker, B. (2012). "lme4: Linear mixed-effects models using Eigen and Eigen++ [computer program]." www.R-project.org/package=lme4 (Last viewed 8/1/2021).

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., and Shinn-Cunningham, B. G. (2014). "Cochlear neuropathy and the coding of supra-threshold sound," *Front. Syst. Neurosci.* **21**, 8–26.

Biberger, T., and Ewert, S. D. (2016). "Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility," *J. Acoust. Soc. Am.* **140**, 1023–1038.

Biberger, T., and Ewert, S. D. (2017). "The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking," *J. Acoust. Soc. Am.* **142**, 1098–1111.

Burgess, A. E., and Colborne, B. (1988). "Visual signal detection. IV. Observer inconsistency," *J. Opt. Soc. Am. A* **5**, 617–627.

Cabrera, L., Varnet, L., Buss, E., Rosen, S., and Lorenzi, C. (2019). "Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity," *J. Acoust. Soc. Am.* **146**, 2415–2429.

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

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906–2919.

Derleth, R. P., Dau, T., and Kollmeier, B. (2001). "Modeling temporal and compressive properties of the normal and impaired auditory system," *Hear. Res.* **159**, 32–149.

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**, 1181–1196.

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

Ewert, S. D., Paragouty, N., and Lorenzi, C. (2018). "A two-path model of auditory modulation detection using temporal fine structure and envelope cues," *Eur. J. Neurosci.* **51**, 1265–1278.

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). "Noise in the nervous system," *Nat. Rev. Neurosci.* **9**, 292–303.

Furman, A. C., Kujawa, S. G., and Liberman, M. C. (2013). "Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates," *J. Neurophysiol.* **110**, 577–586.

García-Pérez, M. A. (1998). "Forced-choice staircases with fixed step sizes: Asymptotic and small-sample properties," *Vis. Res.* **38**, 1861–1881.

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

- Goodman, D. F. M., Winter, I. M., Léger, A. C., de Cheveigné, A., and Lorenzi, C. (2018). "Modelling firing regularity in the ventral cochlear nucleus: Mechanisms, and effects of stimulus level and synaptopathy," *Hear. Res.* **358**, 98–110.
- Goris, R. L., Movshon, J. A., and Simoncelli, E. P. (2014). "Partitioning neuronal Variability," *Nat. Neurosci.* **17**, 858–865.
- Green, D. M. (1964). "Consistency of auditory detection judgments," *Psychol. Rev.* **71**, 392–407.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*, Vol. 1 (Wiley, New York).
- Hasan, B. A. S., Joosten, E., and Neri, P. (2012). "Estimation of internal noise using double passes: Does it matter how the second pass is delivered?," *Vis. Res.* **69**, 1–9.
- Ives, D. T., Kalluri, S., Strelcyk, O., Sheft, S., Miermont, F., Coez, A., Bizaguete, E., and Lorenzi, C. (2014). "Effects of noise reduction on AM perception for hearing-impaired listeners," *J. Assoc. Res. Otolaryngol.* **15**, 839–848.
- Javel, E., and Viemeister, N. F. (2000). "Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination," *J. Acoust. Soc. Am.* **107**, 908–921.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Am.* **124**, 422–438.
- Jones, P. R., Moore, D. R., Amitay, S., and Shub, D. E. (2013). "Reduction of internal noise in auditory perceptual learning," *J. Acoust. Soc. Am.* **133**, 970–981.
- Jørgensen, S., Ewert, S. D., and Dau, T. (2013). "A multi-resolution envelope-power based model for speech intelligibility," *J. Acoust. Soc. Am.* **134**, 436–446.
- King, A., Varnet, L., and Lorenzi, C. (2019). "Accounting for masking of frequency modulation by amplitude modulation with the modulation filter-bank concept," *J. Acoust. Soc. Am.* **145**, 2277–2293.
- Lopez-Poveda, E. A. (2014). "Why do I hear but not understand? Stochastic undersampling as a model of degraded neural encoding of speech," *Front. Neurosci.* **30**(8), 348.
- Lorenzi, C., Soares, C., and Vonner, T. (2001). "Second-order temporal modulation transfer functions," *J. Acoust. Soc. Am.* **110**, 1030–1038.
- Lu, Z. L., Chu, W., and Doshier, B. A. (2006). "Perceptual learning of motion direction discrimination in fovea: Separable mechanisms," *Vis. Res.* **46**, 2315–2327.
- Lu, Z. L., and Doshier, B. A. (2004). "Perceptual learning retunes the perceptual template in foveal orientation identification," *J. Vis.* **4**, 44–56.
- Lu, Z. L., and Doshier, B. A. (2008). "Characterizing observers using external noise and observer models: Assessing internal representations with external noise," *Psychol. Rev.* **115**, 44–82.
- Marmel, F., Rodríguez-Mendoza, M. A., and Lopez-Poveda, E. A. (2015). "Stochastic undersampling steepens auditory threshold/duration functions: Implications for understanding auditory deafferentation and aging," *Front. Aging Neurosci.* **7**, 63–75.
- McDermott, J. H., and Simoncelli, E. P. (2011). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," *Neuron* **71**, 926–940.
- Moore, B. C., Füllgrabe, C., and Şek, A. (2009). "Estimation of the center frequency of the highest modulation filter," *J. Acoust. Soc. Am.* **125**, 1075–1081.
- Moore, B. C., and Şek, A. (1992). "Detection of combined frequency and amplitude modulation," *J. Acoust. Soc. Am.* **92**, 3119–3131.
- Moore, B. C. J., and Şek, A. P. (1994). "Discrimination of modulation type (amplitude modulation or frequency modulation) with and without background noise," *J. Acoust. Soc. Am.* **96**, 726–732.
- Moore, B. C. J., and Şek, A. P. (1996). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," *J. Acoust. Soc. Am.* **100**, 2320–2331.
- Moore, B. C. J., and Şek, A. (2019). "Discrimination of the phase of amplitude modulation applied to different carriers: Effects of modulation rate and modulation depth for young and older subjects," *J. Acoust. Soc. Am.* **146**, 1696–1704.
- Moore, B. C. J., Şek, A., Vinay, and Füllgrabe, C. (2019). "Envelope regularity discrimination," *J. Acoust. Soc. Am.* **145**, 2861–2870.
- Moore, B. C. J., Mariathasan, S., and Şek, A. P. (2019). "Effects of Age and Hearing Loss on the Discrimination of Amplitude and Frequency Modulation for 2- and 10-Hz Rates," *Trends Hear.* **23**, 233121651985396.
- Neri, P. (2010). "How inherently noisy is human sensory processing?," *Psych. Bull. Rev.* **17**, 802–808.
- Neri, P. (2013). "The statistical distribution of noisy transmission in human sensors," *J. Neural Eng.* **10**, 016014.
- 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**, 1–15.
- Paraouty, N., Ewert, S. D., 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**, 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**, 4123–4137.
- Pelli, D. G., and Blakemore, C. (1990). "The quantum efficiency of vision," in *Vision: Coding and Efficiency*, edited by Colin Blakemore (Cambridge University Press, Cambridge, 1990), pp. 3–24.
- Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., and Polley, D. B. (2020). "Bottom-up and top-down neural signatures of disordered multi-talker speech perception in adults with normal hearing," *eLife* **9**, e51419.
- R Core Team (2018). "R: A language and environment for statistical computing," <https://www.R-project.org> (Last visit 8/1/2021).
- Saberi, K., and Hafter, E. R. (1995). "A common neural code for frequency-and amplitude-modulated sounds," *Nature* **374**, 537–539.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Singh, N. C., and Theunissen, F. E. (2003). "Modulation spectra of natural sounds and ethologic theories of auditory processing," *J. Acoust. Soc. Am.* **114**, 3394–3411.
- Spiegel, M. F., and Green, D. M. (1981). "Two procedures for estimating internal noise," *J. Acoust. Soc. Am.* **70**, 69–73.
- Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 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**, 2040–2050.
- Thoret, E., Varnet, L., Boubenec, Y., Fériere, R., Le Tourneau, F. M., Krause, B., and Lorenzi, C. (2020). "Characterizing amplitude and frequency modulation cues in natural soundscapes: A pilot study on four habitats of a biosphere reserve," *J. Acoust. Soc. Am.* **147**, 3260–3274.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Vilidaitė, G., and Baker, D. H. (2017). "Individual differences in internal noise are consistent across two measurement techniques," *Vis. Res.* **141**, 30–39.
- Wallaert, N., Moore, B. C., 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.
- Wallaert, N., Varnet, L., Moore, B. C., and Lorenzi, C. (2018). "Sensorineural hearing loss impairs sensitivity but spares temporal integration for detection of frequency modulation," *J. Acoust. Soc. Am.* **144**, 720–733.
- 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**, 3093–3104.
- 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., Kreft, H. A., and Oxenham, A. J. (2020). "The role of cochlear place coding in the perception of frequency modulation," *eLife* **9**, e58468.
- Wright, B. A., and Dai, H. (1998). "Detection of sinusoidal amplitude modulation at unexpected rates," *J. Acoust. Soc. Am.* **104**, 2991–2996.
- Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargava, A., and Cao, K. (2005). "Speech recognition with amplitude and frequency modulations," *Proc. Nat. Acad. Sci. U.S.A.* **102**, 2293–2298.
- Zwicker, E. (1952). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenz-modulation eines Tones" ("The limits of audibility of amplitude modulation and frequency modulation of a tone"), *Acta Acust. united Ac.* **2**, 125–133.