Speeded detection of spectrotemporal modulations of sound

Lidwien C.E. Veugen[a], A. John van Opstal[a], and Marc M. van Wanrooij[a,1]

[a] Department of Biophysics, Donders Institute for Brain, Cognition and Behavior, Radboud University

[1] To whom correspondence should be addressed. E-mail: marc.vanwanrooij@donders.ru.nl

We tested whether joint spectrotemporal sensitivity follows from spectrotemporal separability for normal-hearing conditions and for impaired-hearing simulations. In a manual reaction-time task, normal-hearing listeners had to detect the onset of a ripple (with density between 0-8 cycles/octave and a fixed modulation depth of 50%), that moved up or down the log-frequency axis at constant velocity (between 0-64 Hz), in an otherwise-unmodulated broadband white-noise. Spectral and temporal modulations elicited band-pass filtered sensitivity characteristics, with fastest detection rates around 1 cycle/oct and 32 Hz for normal-hearing conditions. These results closely resemble data from other studies that typically used the modulation-depth threshold as a sensitivity measure for spectral-temporal modulations. To simulate hearing-impairment, stimuli were processed with a 6-channel cochlear-implant vocoder, and a hearing-aid simulation that introduced spectral smearing and low-pass filtering. Reaction times were always much slower compared to normal hearing, especially for the highest spectral densities. Binaural performance was predicted well by the benchmark race model of statistical facilitation of independent monaural channels. For the impaired-hearing simulations this implied a "best-of-both-worlds" principle in which the listeners relied on the hearing-aid ear to detect spectral modulations, and on the cochlear-implant ear for temporal-modulation detection. Although singular-value decomposition indicated that the joint spectrotemporal sensitivity matrix could be largely reconstructed from independent temporal and spectral sensitivity functions, in line with time-spectrum separability, a significant inseparable spectral-temporal interaction was present in all hearing conditions. These results imply that the reaction-time task yields a solid and effective objective measure of acoustic spectrotemporal modulation sensitivity, which may also be applicable to hearing-impaired individuals.

ripple | binaural | bimodal | vocoder | modulation transfer function | reaction time

Introduction

Human speech and other complex sounds in the natural environment are typically dynamic signals that rapidly change in amplitude over both time and frequency. Fluctuations in the temporal domain provide information about the rhythm of speech, such as syllable and word boundaries, whereas variations in the spectral domain are essential for formant and voice pitch perception (Liberman 1996). Sensitivity to these joint spectral and temporal modulations is deemed crucial for the identification of complex sound features (McDermott and Simoncelli 2011) and for speech comprehension (Elliott and Theunissen 2009; Shannon et al. 1995).

Spectrotemporal dynamic ripples have been introduced in psychoacoustics to investigate the spectrotemporal modulation sensitivity of auditory perception. Ripples are broadband noise stimuli that are modulated sinusoidally in amplitude over time and/or frequency (Bernstein and Green 1998; Supin et al. 1994). Ripples are ideal to assess hearing performance as they represent, but are not recognizable as, naturalistic sounds. Sensitivity of the healthy human auditory system has been studied thoroughly with ripples and generally shows a band- or low-pass response to spectral and temporal modulations, reflecting the limits of auditory sensitivity at higher modulation rates (Chi et al. 1999; Viemeister 1998; Zheng et al. 2017). Speech understanding is thought to relate mostly to joint spectrotemporal sensitivity. Chi et al. (1999) reported that in normal-hearing listeners the modulation transfer function of combined spectrotemporal ripples is highly separable, as it can be well approximated by the product of a single temporal and spectral filter. Separability implies that the joint spectrotemporal sensitivity can be directly obtained from pure temporal and spectral sensitivity measurements.

In the present study we used manual reaction times to construct the spectrotemporal modulation transfer function (stMTF), rather than the conventionally used modulation detection or discrimination thresholds. Research in monkeys shows that reaction times systematically depend on acoustic modulation rates (Massoudi et al. 2014). Several models have been proposed to explain the underlying process of response latency in reaction-time tasks (Ratcliff and Van Dongen 2011). It is commonly assumed that a decision signal rises with accumulating evidence of the stimulus, until a certain threshold is reached that triggers the response. As such, reaction times are directly related to the difficulty of a task and could thus provide more detailed information on the audibility of spectrotemporal ripples.

Furthermore, reaction times allow for testing the presence or absence of binaural integration based on monaural responses, by comparing binaural reaction times against the prediction of a so-called 'race model'. In such a model, the signals from either ear compete independently to reach the detection threshold, so that the response latency is determined by the winner of an independent parallel race between the two ears (Raab 1962). Due to statistical facilitation, this race to threshold leads to faster reaction times for binaural stimulation than for monaural stimuli, as the distribution of minimum monaural reaction times yields faster responses than those produced by either ear (Gielen et al. 1983; Hershenson 1962). However, when this so-called redundant stimulus effect differs from the race-model prediction, it could imply true binaural integration in an underlying neural interaction process (Gielen et al. 1983; Miller 1982; Schröter et al. 2007).

We tested whether reaction times are a solid objective measure of auditory sensitivity to moving ripples with various spectrotemporal modulations for normal-hearing listeners. We assessed (in)separability of joint spectrotemporal sensitivity and investigated how binaural listening affected modulation sensitivity compared to monaural listening conditions by comparing the data with the race-model prediction. As a validation of our reaction-time paradigm, we also collected data under more challenging impaired-hearing simulations that are known to affect temporal and spectral sensitivity (Bacon and Viemeister 1985;

Golub et al. 2012; Henry et al. 2005; Moore and Glasberg 2001) and binaural integration (Ausili et al. 2019; Sharma et al. 2019, 2021; Veugen et al. 2016b).

Methods

Listeners. Six listeners participated in this study (3 male, ages 20-25 years), none of whom reported a history of auditory deficits. All listeners had normal hearing (< 20 dB HL) in both ears from 125 to 8000 Hz. Except for two of the authors, listeners were naïve to the purpose of the experiments. The study was approved by the Local Ethics Committee of the Radboud University Nijmegen (protocol number 40327.091.12).

Apparatus. Listeners were seated in an acoustically shielded sound chamber. Stimuli were presented through TDH 39 headphones (Telephonics Corporation, Farmingdale, NY, USA). For sound processing and data acquisition we used Tucker Davis Technologies System 3 (Alachua, FL, USA). Stored sounds were sent via the PC to a real-time processor (RP2.1) at a sampling rate of 48828.125 Hz, and passed through a programmable attenuator (PA5). Stimuli were set at a comfortable, well-audible loudness of 65 dB(A) (calibrated using a KEMAR head calibration set, connected to a Brüel & Kjaer measuring amplifier type 2610 [Nærum, Denmark]).

Stimuli. Dynamic ripples were created in MATLAB [version R2012a; Mathworks Inc., Natick, MA, USA] as described by Depireux, Simon, Klein, and Shamma (2001). The carrier of these stimuli consisted of a broadband spectrum of multiple harmonic tones, each described by:

$$c_i(t) = \sin(2\pi f_i t + \varphi_i) \qquad [1]$$

where t is time (s), f_i frequency (Hz) of the i-th harmonic, and φ_i is its phase (rad). In this experiment, the broadband carrier consisted of 128 harmonic tones, equally spaced (20 tones/octave) over 6.4 octaves (250 Hz-20.4 kHz). All components had random phase except for the first (φ_0=0). The i-th frequency was determined by $f_i = f_0 2^{i/20}$ with f_0 = 250 Hz the lowest frequency, and i = 0 - 127.

The spectrotemporal envelope determined the ripple fluctuations in amplitude over time and/or frequency:

$$s_i(t, x_i) = 1 + \Delta M \cdot \sin(2\pi\omega t + 2\pi\,\Omega\, x_i) \qquad [2]$$

with t is time (s), x_i is the position on the frequency axis (in octaves above the lowest frequency), ΔM is the modulation depth (here set to 0.5 for all components), ω is the ripple velocity (Hz) and Ω is ripple density (cycles/octave).
Together the carrier and the modulator formed the dynamic ripple in our experiments as follows:

$$R(t) = \begin{cases} t < t_{onset}, & \sum_i c_i(t) \\ t > t_{onset}, & \sum_i s_i(t - t_{onset}, x_i) \cdot c_i(t) \end{cases} \qquad [3]$$

The modulated sounds were thus preceded by a non-modulated harmonic complex ($c_i(t)$) with a randomized duration (t_{onset}) between 700 and 1200 ms with a step size of 100 ms. Moving ripples were presented with velocities of 0 Hz and ±[0.5; 1; 2; 4; 8; 16; 32; 64] Hz and densities [0; 0.25; 0.5; 0.75; 1; 2; 4; 8] cycles/octave, yielding a total of (17 velocities x 8 densities =) 136 different stimuli.

Cochlear implant simulation. cochlear-implant vocoder simulations were created using a previously described method by Litvak et al. (Litvak et al. 2007) that models the Advanced Bionics Harmony cochlear-implant processor. Briefly, the vocoder algorithm works as follows. After resampling the input signal to 17.4 kHz, the vocoder applied a high-pass pre-emphasis filter (cut-off at 1.5 kHz). Then, the signal was band-pass filtered by a short-time Fourier transform with 256 bins and 75% temporal overlap (192 bins). Bins were grouped into 6 non-overlapping logarithmically spaced channels (Fig. 1B; at center frequencies: 452, 715, 1132, 1792, 2836 and 4488 Hz). Random-phase noise bands with similar center frequencies were modulated with amplitudes equal to the square root of the total energy in the channel. The channels were summed, and inverse short-time Fourier transformed to reproduce a temporal waveform for presentation to the listeners.

We used 6 vocoder channels to simulate hearing via a cochlear implant, as cochlear-implant users are typically unable to effectively utilize information from all available cochlear-implant channels (Henry and Turner 2003). Normal-hearing listeners have shown similar performance as cochlear-implant users for speech understanding scores in quiet with 4-6 channels (Loizou et al. 1999). This is in line with pilot experiments in our lab that revealed that five normal-hearing listeners achieved a performance level of ~80% in a consonant-vowel-consonant recognition test, when the words were vocoded with only 6 channels.

Hearing aid simulation. hearing-aid simulations were generated by using a fourth-order Butterworth low-pass filter with a cut-off at 500 Hz, mimicking residual hearing in the lower frequencies present in the bimodal users of our previous studies (Sharma et al. 2019; Veugen et al. 2016a). Additionally, the loss of frequency selectivity (spectral smearing) was simulated as previously described by Baer and Moore (1994). Asymmetrically broadened auditory filters were used with broadening factors 6 and 3 for the lower and upper branch respectively, as these are representative for moderate-to-severe hearing impairment (Glasberg and Moore 1986). The cochlear implant and hearing-aid simulated stimuli were normalized to the same root-mean-squared value as the original non-vocoded sounds (Fig. 1 visualizes the effect of the cochlear implant and hearing-aid simulation on ripples).

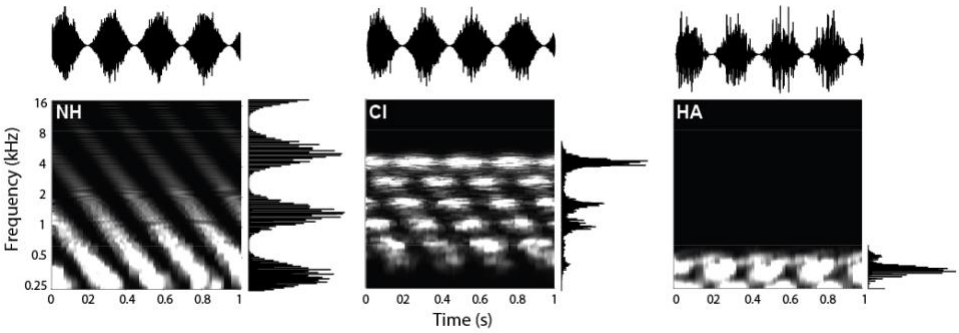

Fig. 1. Moving-ripple spectrograms. Ripple with velocity 4 Hz and density 0.5 cycles/octave) for the normal-hearing condition (NH), cochlear implant simulation (CI) and hearing-aid simulation (HA). The signals on the top row represent the temporal waveforms for a purely amplitude-modulated sound (4 Hz, 0 cycles/octave). The signal on the right of each Fig. visualizes a pure stationary spectral ripple modulation (0 Hz, 0.5 cycles/octave). For clarity, the sound is shown after t_{onset} (at t=0).

Paradigm. Listeners were instructed to press a button as quickly as possible when they heard the sound change from static noise to modulated ripple. Modulated ripples lasted for 3000 ms, unless the button was pressed, in which case the sound was ended prematurely, and the next trial was initiated after a brief (0.5 - 1 s) period of silence between each trial. If the button was pressed before ripple onset, the trial was reiterated, but no more than 4 times. The outcome measure of the experiment was the listener's manual reaction time, which was defined as the time between the onset of the ripple and the moment the button was pressed.

We tested five different listening conditions; acoustic stimuli were presented 1) monaurally (monaural normal hearing), 2) monaurally via cochlear-implant vocoder (unimodal cochlear implant), 3) monaurally via hearing-aid simulation (unimodal hearing aid), 4) binaurally (binaural normal hearing), 5) binaurally via bimodal cochlear implant and hearing-aid simulation (bimodal). In monaural conditions, both ears were tested separately. In the bimodal condition, cochlear implant and hearing aid were tested in both the right and left ear in different sessions. We did not test the binaural unimodal listening conditions (cochlear implant-cochlear implant or hearing aid-hearing aid). Each stimulus was presented 5 times in each listening condition. A complete data set thus contained a total of 6120 stimuli, which were split in 12 sessions of 30-40 minutes, each containing 510 trials. Sessions were distributed over 6 days of two sessions each. Ripples and conditions were presented in pseudo-randomized order.

Because of time constraints, data collection was not fully completed in the four naïve listeners. Two naïve listeners completed 11 out of 12 sessions, with all ripples measured at least twice. The other two listeners completed 9 out of 12 sessions (20 and 23 ripples not measured, respectively; 83 and 141 ripples were presented only once in these listeners).
The four naïve listeners performed one training session under normal hearing conditions prior to the recording sessions, to become familiarized with the ripple stimuli and experimental procedures. We observed no systematic change in the average reaction times during the training session for these four listeners, or over the time course of the experimental sessions for all six listeners. This observation indicates that procedural learning effects did not confound the reaction-time data.

Analysis. Data analysis was performed with custom-written MATLAB software. Reaction times generally show a skewed distribution with an extended tail towards longer reaction times. To obtain normally distributed data (Carpenter et al. 2009), the reaction-time data were transformed to their reciprocal (1/reaction time), referred to here as 'promptness' (1/s). This also allows the measurements to be more readily interpreted as sensitivity measures to the different spectrotemporal modulations, as a higher/lower promptness (as opposed to a shorter/longer reaction time) indicates a higher/lower sensitivity. Responses were pooled across listeners and ears for grand average analyses. Reaction times below 150 ms (clear anticipatory responses) were removed from the analysis. If a response was not initiated within 3 s (considered a sign for inattentiveness, or of an inability to detect the ripple), we set the response time (promptness) to 3 s (1/3 s^{-1}). Non-responses were found in 10% of the trials under normal-hearing conditions and in 46% of the trials of the hearing-impaired conditions (especially at the high spectral modulations). We do not explicitly account for the percentage of non-responses but note that in our analyses a higher number of non-responses would yield a median promptness (reaction time) closer to 1/3 s^{-1} (3 s). The non-modulated sound (velocity 0 Hz and density 0 cycles/octave) served as a catch stimulus, to determine the guess (or false-alarm) rate of the participant. The guess rate varied from 7% for binaural normal hearing, to 21% for monaural hearing-aid listening, with the average guess rate across conditions at 12%.

Spectrotemporal transfer function. For each of the five listening conditions (monaural and binaural normal-hearing, and monaural and bimodal cochlear implant and hearing-aid simulation), we calculated the mean promptness per ripple to construct a two-dimensional spectrotemporal modulation transfer function $stMTF(\omega, \Omega)$ as a joint function of ripple density, Ω, and ripple velocity, ω. Similarly, we determined the temporal modulation transfer function tMTF, $F(\omega)$, and the spectral modulation transfer function sMTF, $G(\Omega)$, for the 0-density and 0-velocity stimuli, respectively.

Separability. To analyze the degree of separability of the stMTF, we applied singular value decomposition (SVD) for all listening conditions. SVD transforms the stMTF into two unitary matrices containing temporal and spectral singular vectors, respectively, and a rectangular diagonal matrix that contains the singular values: $stMTF(\omega,\Omega)=F(\omega)\cdot\Sigma\cdot G(\Omega)$. In case of a fully separable stMTF, the spectral and temporal components are independent of each other and the total of all 136 spectrotemporal responses can be expressed by the vectorial outer product of a single temporal $F_1(\omega)$ (17 components) and spectral $G_1(\Omega)$ (8 components) modulation transfer function, as follows:

$$stMTF_1(\omega, \Omega) = \sigma_1 \cdot F_1(\omega) \times G_1(\Omega) \quad [4]$$

with σ_1 the largest singular value.
We calculated the separability index (cf. inseparability index, as used by, for example, Massoudi et al. 2015; Versnel et al. 2009), which ranges between zero (totally inseparable) to one (fully separable), and is based on the relative dominance of the first SVD component:

$$\alpha_1 = \frac{\sigma_1^2}{\sum_i^n \sigma_i^2} \quad [5]$$

The separable stMTF estimate was reconstructed according to Eq. 4.

We also reconstructed the stMTF based on the first two singular values, according to

$$stMTF_{12}(\omega, \Omega) = \sigma_1 \cdot F_1(\omega) G_1(\Omega) + \sigma_2 \cdot F_2(\omega) G_2(\Omega) \quad [6]$$

and determined the relative contribution of the first two SVD components:

$$\alpha_2 = \frac{\sigma_1^2 + \sigma_2^2}{\sum_i^n \sigma_i^2} \quad [7]$$

Before applying the SVD, the stMTF data was centered by subtracting the mean promptness for each listening condition. This mean was added to the reconstructions.

Race model. We compared the observed reaction times for binaural hearing with the quantitative predictions of performance based on the monaural reaction times, using the race model of statistical facilitation. This model assumes independence of the two monaural processes (Gielen et al. 1983; Raab 1962). Any violation to the race model suggests neural interactions when processing the input from both ears:

$$P(\tau_{race} \leq t) = P(\tau_{M1} \leq t) + P(\tau_{M2} \leq t) - P(\tau_{M1} \leq t) \times P(\tau_{M2} \leq t) \quad [8]$$

with $P(\tau \leq t)$ the cumulative probability function (CDF) of an observed reaction time g at time t; M_1 and M_2 represent monaurally-presented stimuli (normal hearing, cochlear implant and hearing aid). We estimated the cumulative probability density functions (CDF) from the promptness values. The race model CDF was constructed from the two monaural CDFs according to Eqn. 8. For comparative purposes, we calculated the difference in the medians (at the 50% cumulative probability level) between actual performance and race-model predictions. Ripples, for which fewer than 10 responses were collected, were discarded from this analysis because no reliable CDF could be constructed. Non-responses (reaction times > 2500 ms) were also discarded from the race-model analysis.

Statistics. Data were always reported as mean values ± 1 standard deviation. We also calculated 95% confidence intervals of promptness for the pure temporal and spectral ripples. As a criterion of significance for a statistical difference we took $p<0.05$.

Results

Reaction-time task. We will first illustrate the systematic dependence of the manual reaction times on the acoustic conditions with the data of one listener (Fig. 2). Pure amplitude-modulated noises elicit cumulative distributions of reaction times in the binaural listening condition that shift systematically with the velocity of the stimulus (Fig. 2A). The cumulative distributions are plotted on a probit scale as a function of the reciprocal of reaction times (Carpenter and Williams 1995; Corneil et al. 2002). In this format, the data points for each velocity fall closely on a straight line indicating that the promptness responses form a normal distribution. One may note that the singly-most distinguishing feature of these lines is that they are shifted versions of each other, with the slopes being similar across the velocity modulations. This may suggest that the mean reaction time and its inverse, promptness (the promptness at 50% cumulative probability) is a good point estimate of the effect of a ripple's velocity on reaction speed. Indeed, the mean promptness of this listener systematically and monotonically increases when velocity is increased (Fig. 2D). Also, for spectral modulations, reaction time distributions (Fig. 2B) and mean promptness (Fig. 2E) vary systematically with a ripple's density, albeit that reaction speed seems to decrease with increasing density. Listening condition also affected this listener's response speed, as exemplified for a [1 Hz, 0 cycles/octave] modulation (Figs. 2C, F), with binaural hearing eliciting the fastest responses and the monaural hearing-aid simulation yielding the slowest responses. In the following sections, we will quantify this relationship with promptness and the ripple modulation parameters for all listeners.

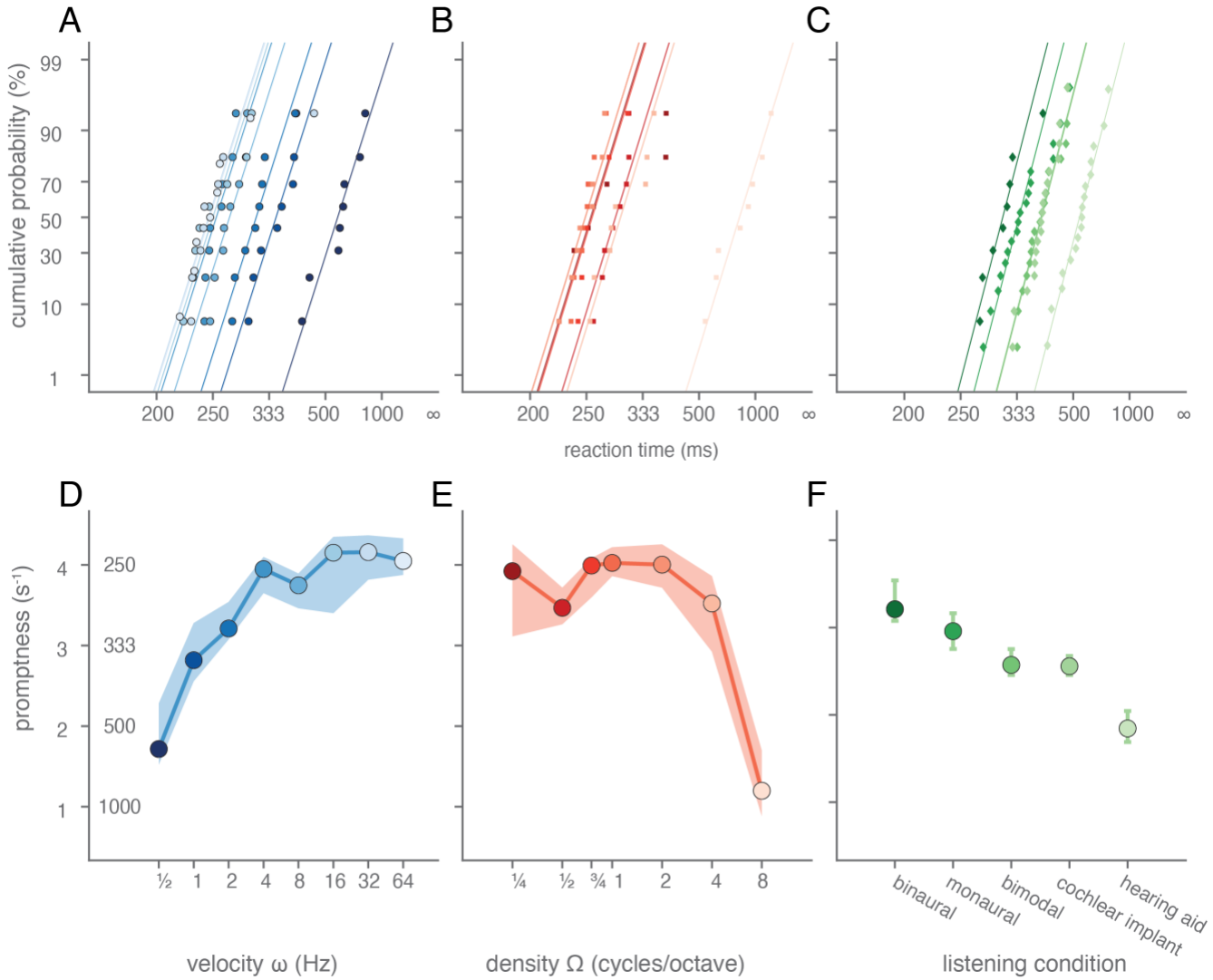

Fig. 2. Example reaction times to temporally and spectrally modulated target sounds. A-C) Reaction time distributions plotted in the form of cumulative percentage probability, on a probit scale, as a function of reciprocal reaction time (promptness), measured in one listener (L1) in various listening regimes. The temporal and spectral modulation of the target sound and the listening condition was varied, leading to changes in the listener's detection speed. Each curve corresponds to a different modulation or listening condition. A) Reaction time distributions to amplitude modulations (density = 0 cycles/octave) in a binaural listening condition are shown. As each distribution seemed to lie close to a straight line, regression lines are drawn, for illustrative purposes. Reaction time distributions to B) spectral modulations (velocity = 0 Hz) in a binaural listening condition and to C) different listening conditions for a single amplitude-modulated target sound (velocity = 2 Hz, density=0 cycles/octave) are shown. D-F) Mean promptness (cf. reaction time, see left ordinate) is plotted as a function of D) velocity, E) density, and F) listening condition (stimulus parameters are the same as in A-C). Cochlear implant and hearing aid indicate cochlear-implant and hearing-aid simulation, respectively. Colours in each plot indicate velocity, density and listening condition as indicated by the abscissa-values of the coloured, filled circles. Coloured patches indicate 95% confidence intervals of the mean.

Temporal-only modulations. We will first elaborate on how reaction times reflect the detection of temporal-only modulations (Fig. 3A). For the normal-hearing listening conditions (binaural and monaural), the mean promptness as a function of velocity for the purely temporal amplitude modulations (density = 0 cycles/octave) resembled a high-pass characteristic (Fig. 3A, dark blue circles and light blue triangles). Responses were fastest (higher promptness) for the highest absolute velocities and were slower for lower velocities. If the sounds were vocoded, simulating hearing device processing (cochlear implant, hearing aid, bimodal), the promptness dropped especially for the higher velocities, so that the curve exhibited bandpass properties. Responses were now fastest for intermediate absolute velocities and were slower for both higher and lower velocities. Overall, both the maximal promptness and at which velocity this was attained were affected by listening condition; the fastest responses, with an average promptness of 3.3 (monaural, light blue triangles), 3.5 (binaural, dark blue circles), 2.5 (bimodal, pink squares), 2.5 (cochlear implant, light green diamonds) and 2 s^{-1} (hearing aid, dark green triangles) were observed at ±32, ±32, ±16, ±16, ±8 Hz, respectively. The longer response times to pure amplitude modulations under impaired-hearing conditions clearly implicate increased difficulty in the detection of temporal modulations. Differences in response times between binaural and monaural listening will be considered in more detail below with race-model predictions. Responses to upward (<0 Hz) and downward (>0 Hz) moving ripples were very similar; correlation coefficients between the responses to up- and downward ripples were between 0.91-0.99 for all listening conditions.

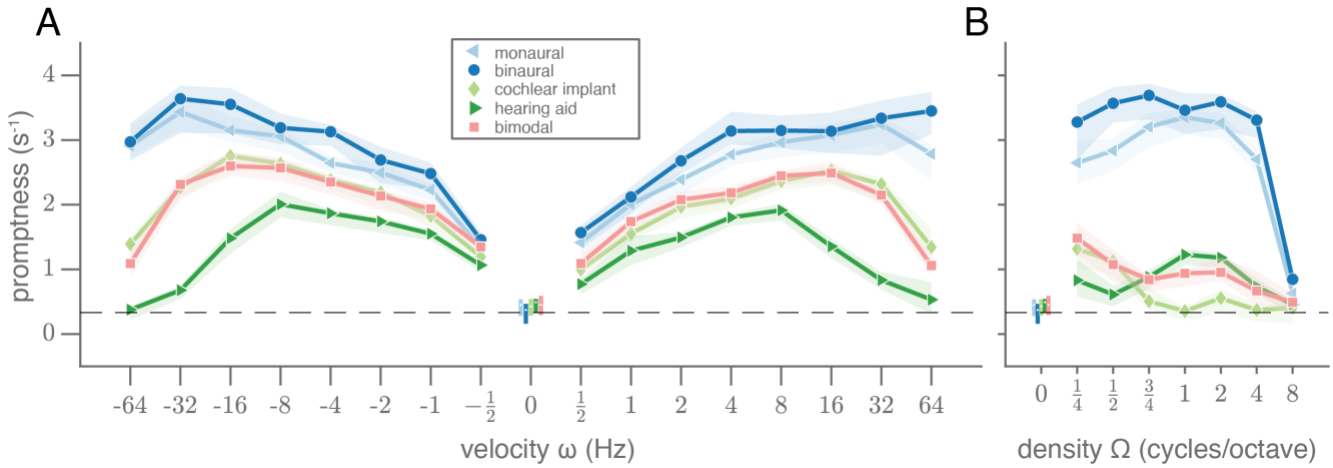

Fig. 3. Pure temporal and spectral modulation transfer functions. A) Mean promptness (line and markers) and 95% confidence interval of the mean across subjects (patch) as a function of velocity (density = 0 cycles/octave) and B) as a function of density (velocity = 0 Hz), for each listening condition: normal-hearing monaural and binaural conditions, and the impaired-hearing cochlear-implant, hearing-aid and bimodal simulations. White dots and colored vertical lines around 0 indicate the mean promptness and its 95% confidence interval for the catch trials with unmodulated stimuli ([velocity, density] = [0 Hz, 0 cycles/octave]).

Spectral-only modulations. For the static ripples (purely spectral modulations at velocity = 0 Hz), the promptness as a function of density resembled a low-pass characteristic, at least for the binaural and monaural normal-hearing conditions (Fig. 3B, dark blue circles and light blue triangles, respectively). For these conditions, detection is very poor for the highest density of 8 cycles/octave. This property presumably reflects the limits of resolvable power of the human auditory filters, leading to a poorer detection of spectral patterns with finer spectral detail.

Responses made for the cochlear-implant simulation (light green diamonds) resembled a band-pass filter characteristic with a cutoff around 0.75 cycles/octave and responses in the hearing-aid condition (dark green triangles) followed a band-pass characteristic with its highest promptness around 1-2 cycles/octave. Bimodal responses (pink squares) resembled the best values of the cochlear-implant and hearing-aid conditions. Overall, the impact of hearing-impairment simulation on reaction times was generally larger for the spectral modulations (Fig. 3B) than for the temporal modulations (Fig. 3A). This behavioral finding seems in line with the acoustic effects of the vocoders on the sounds, that preserve temporal modulations to some extent (cf. Fig. 1, top row), but heavily perturb spectral modulations (cf. Fig. 1, side columns).

Joint spectrotemporal modulation. Figs. 4A-E show the spectrotemporal modulation transfer functions for the two normal-hearing conditions, and for the three impaired-hearing simulations, for all joint spectrotemporal ripples, as mean promptness (averaged across listeners) per ripple density (abscissa) and velocity (ordinate). Deep red colors correspond to high spectral-temporal sensitivity, dark blue colors to low sensitivity (low promptness values). The results for pure amplitude-modulated stimuli (cf. Fig. 3A) are at the bottom row of the stMTF matrix, at Ω=0 cycles/octave; the results for pure spectral modulations (cf. Fig. 3B) are found along the central column, at ω=0 Hz.

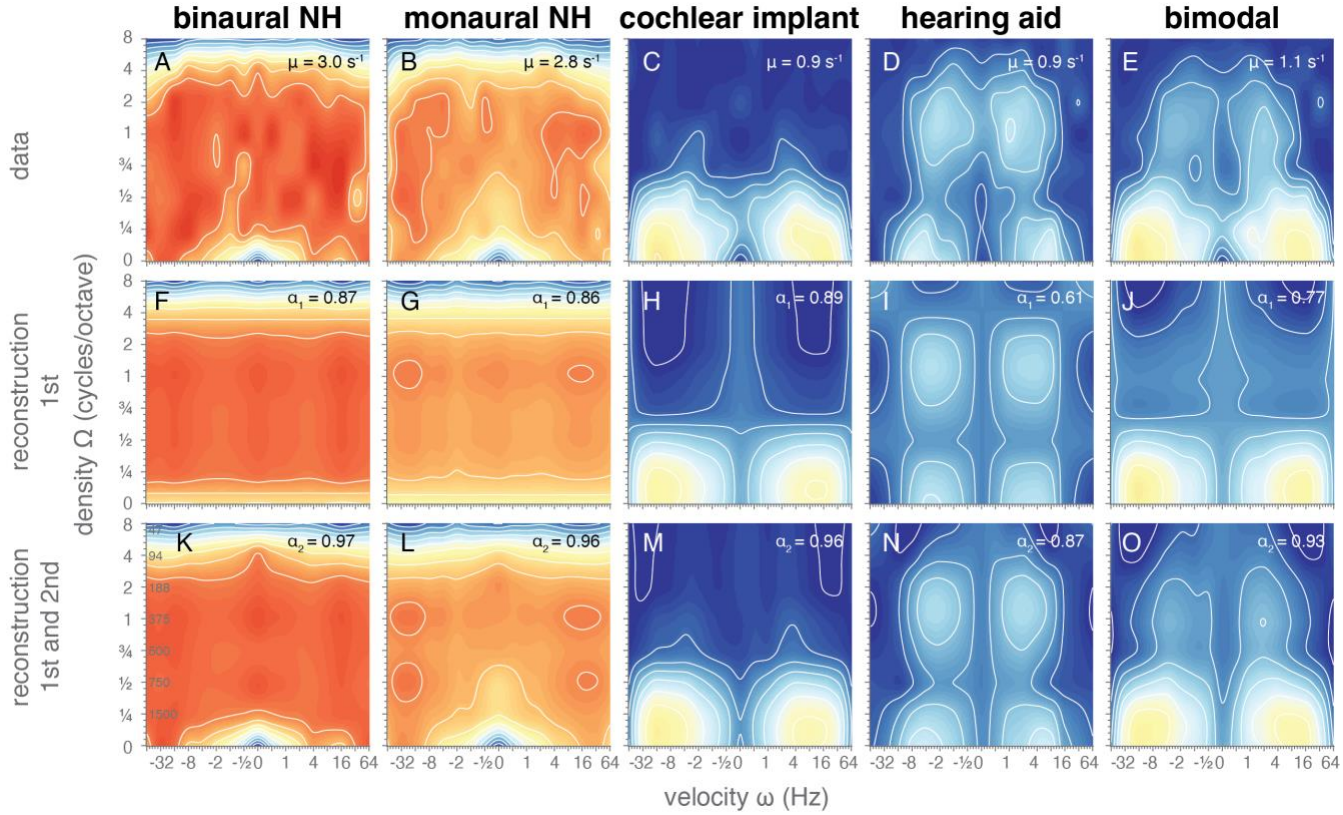

Fig. 4. Spectrotemporal modulation transfer functions. Mean promptness as a function of velocity and density, representing spectrotemporal sensitivity in the normal hearing and simulated hearing-impaired conditions. F-J) Reconstructions of the spectrotemporal modulation transfer functions using only the first singular value after singular value decomposition (SVD; Eq. 4). K-O) Reconstructions of the spectrotemporal modulation transfer functions using the first and second SVD components (Eq. 6).

Ripple detection in the binaural normal-hearing condition (Fig. 4A) was faster (mean promptness = 3.0 s^{-1}) than in the monaural condition (Fig. 4B; mean promptness: 2.8 s^{-1}). Combined spectrotemporal modulation sensitivity for both listening conditions resembled a two-dimensional band-pass for both density and velocity, with fastest reaction times around velocities of 8-16 Hz and densities around 0.75-1 cycles/octave (binaural maximum promptness = 4.0 s^{-1} at [ω, Ω]=[8 Hz, 0.75 cycles/octave]; monaural maximum promptness = 3.7 s^{-1} at [ω, Ω]=[-16 Hz, 1 cycles/octave]). The ripples were well detectable up to and including 4 cycles/octave.

The spectrotemporal modulation transfer functions for the impaired-hearing simulations (Fig. 4C-E) were distinctly slower when compared to normal hearing (mean promptness = 0.9, 0.9 and 1.1 s^{-1} for cochlear implant, hearing aid, and bimodal vocoder, respectively). Temporal modulation sensitivity again showed a band-pass filter characteristic with fastest detection rates around 16 Hz. Cochlear-implant simulations (Fig. 4C) mainly differentially affected the detection of spectral modulations, which is consistent with the modus operandi of a cochlear implant (and vocoders), whereby its band-pass filtering mechanism reduces spectral modulation sensitivity. The fastest responses (maximum promptness = 2.8 s^{-1}) were elicited to -16 Hz amplitude-modulated sounds (Ω = 0 cycles/octave). Ripple detection with the cochlear-implant vocoder became impossible for densities exceeding 0.75 cycles/octave. Of all hearing conditions, listeners reacted slowest for the hearing-aid simulations (Fig. 4D; maximum promptness = 2.0 s^{-1}). However, in contrast to the cochlear-implant condition, higher densities of up to 4 cycles/octave could still be detected if temporal modulation rates were not too fast (>16 Hz). A local dip in promptness exists for ripples with a density around 0.5 Hz.

The bimodal simulation resembled a conjunction of the cochlear-implant and hearing-aid simulation results, seemingly exhibiting a 'best of both worlds' principle (Corneil et al., 2002) with responses almost as fast as for the best unimodal condition (Fig. 4E; maximum promptness = 2.6 s^{-1} at [ω, Ω]=[-16 Hz, 0 cycles/octave]). For high spectral modulation frequencies, the bimodal condition was comparable to the hearing-aid condition; for low spectral modulations it followed the cochlear-implant condition.

Separability. We assessed the degree of separability of the stMTF into a pure temporal and spectral component through singular value decomposition (SVD) using the separability index α_1 (Eq. 5) and the α_2 index (Eq. 7) to determine the relative contribution of the first (Eq. 4) and of the first two components (Eq. 6). If the α_1 index is close to 1, the MTF is considered to be separable.

The central row of Fig. 4 shows the reconstructed $stMTF_1$ for the various hearing conditions. The first-order reconstruction of the stMTF yielded purely orthogonal patterns in the matrix, resulting from the full-separability assumption. For both normal hearing conditions, the separability index α_1 was high (0.87 and 0.86 for binaural and monaural listening conditions), which suggests that the variability in the normal-hearing stMTFs can be captured quite well with a first-order approximation (Figs. 4F,G) and that the matrix is highly separable (eq. 4). Notably missing in the reconstructions are the slow responses to the slow amplitude-modulated sounds (ω<1 Hz, Ω=0 cycles/octave; cf. Figs. 4A,B and Figs. 4F,G at the bottom

of the images near the center). By adding the second singular value with its spectral and temporal components (bottom row of Fig. 4), the stMTF reconstructions improved considerably: α_2=0.96 (monaural) and 0.97 (binaural). Now, the responses to the amplitude-modulated sounds seemed to be accounted for as well.

The separability index was best for the cochlear-implant simulation (Fig. 4H), which equaled or was better than for the normal-hearing conditions (α_1=0.89). The first-order reconstructions were worse for the hearing-aid and bimodal conditions (α_1= 0.61 and 0.77, respectively), suggesting a considerable inseparable spectrotemporal component to the responses of these two listening conditions. Incorporating the first two SVD components improved reconstructions (Figs. 4M-O): α_2=0.96, 0.87, 0.93.

Race models. To investigate to what extent monaural reaction times can predict binaural performance, we used the race model of statistical facilitation, which postulates independence between ears. As an example, Fig. 5A displays the cumulative reaction time probability of listener L6 for a stimulus with a 2 Hz, 0 cycles/octave modulation, for the monaural and binaural normal-hearing conditions, as well as for the promptness that would be reached based on the race model of statistical facilitation. For this ripple, binaural performance was faster than both the monaural condition and the race model. For the simulated listening conditions (Fig. 5B), the bimodal responses to this ripple were faster than the hearing-aid data and resembled the cochlear-implant data and the race model.

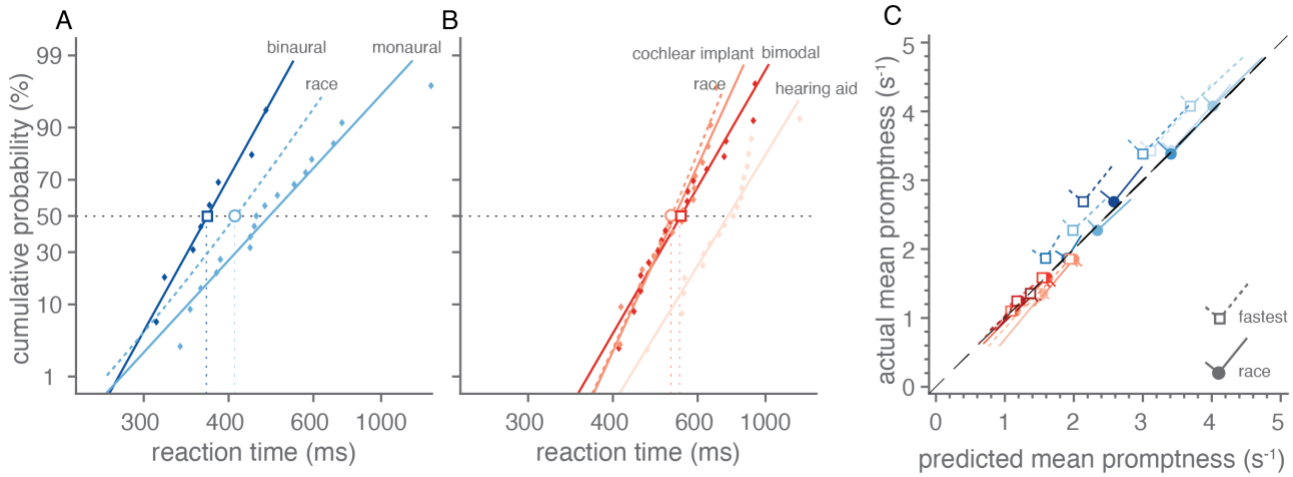

Fig. 5. Binaural versus monaural detection of spectrotemporal modulations. A) Reaction time distributions of listener L6 plotted in the same format as in Figs. 2A-C for a stimulus with a 2 Hz, 0 cycles/octave modulation while listening monaurally (light blue) or binaurally (dark blue). The race model (determined from the monaural distribution according to Eqn. 8) is indicated by a light blue dashed line. Dots indicate individual reaction times; lines indicate the best linear fit through the data. The horizontal dotted line is drawn at a cumulative probability of 0.5 crossing the median reaction times. These medians for the bimodal data and the race model are indicated by an open square and circle, respectively. B) The same as in A, but now for the cochlear-implant, hearing-aid and bimodal listening conditions. Note that the binaural data is slightly faster than the race model, while bimodal data resembles both the cochlear-implant data and the race model. C) The median promptness as observed in the data is plotted as a function of the race-model prediction (closed circles and unbroken lines) and as a function of the fastest-condition prediction (open squares and dashed lines). Blueish and reddish colors indicate binaural and bimodal listening conditions, respectively. For each listener (indicated by different tint), the mean across ripple modulations is indicated by the circles and squares, and standard deviation in the direction of the two axes with largest variability is indicated by lines. Note that the data fall closely on the unity-line (black dashed line) when compared to the race model, which holds both for the mean and the main axis with largest variability. The binaural data are typically faster than the fastest monaural listening condition, while for the bimodal condition there is no clear difference between actual data and both race and fastest model.

Likewise, we compared the predictions from the race model (at the level of the median) and the fastest listening condition with the binaural and bimodal median promptness for all listeners and ripples (Fig. 5C). Overall, binaural performance equaled (Fig. 5C, blue circles on the diagonal) the benchmark race model prediction and was faster than the fastest monaural condition for most ripples (Fig. 5C, blue squares above the diagonal). Bimodal-hearing performance was as fast as both the race model and the fastest monaural condition (Fig. 5C, reddish colors). These results show that, overall, binaural and bimodal performance seemed to follow statistical facilitation (Eqn. 8).

Discussion

Summary. This study used a speeded-response paradigm to determine the auditory spectrotemporal modulation transfer function in human listeners. The reaction times obtained appeared to be a solid and effective objective measure for ripple sensitivity, given its systematic relationship with the parameters that determine both temporal and spectral modulation rates. Sensitivity was highest for ripples with modulations around 16 Hz and 1 cycle/octave and decreased for higher and lower modulation rates. Using simulations of a cochlear-implant, hearing-aid and bimodal restorative hearing, spectrotemporal sensitivity worsened (reaction times increased) compared to normal hearing, in line with the impaired signal processing of the simulations. Although the separability of the stMTF into a spectral and temporal component was high for both the normal-hearing and for the simulated impaired-hearing data, the responses also suggested a significant inseparable second-order spectrotemporal component, with a value between 10-20% of the first component, for all hearing conditions. For all ripples,

binaural and bimodal reaction-time performance was comparable to the prediction of the benchmark race model of statistical facilitation, suggesting independent detection of two monaural signals, rather than true binaural integration.

Constructing the stMTF based on reaction times is a fairly new approach that has been introduced so far only in monkey research (Massoudi et al. 2013, 2014). In those studies, responses to temporal ripples systematically depended on the ripple's period, and we therefore analyzed the pure temporal and spectral modulation sensitivities separately (Fig. 3). All other ripples contained spectral modulations, and therefore had a more abrupt change from stationary noise to the modulated ripple in our stimuli. Like in conventional modulation detection threshold procedures, responses were then driven by the listener's auditory spectrotemporal sensitivity, rather than by the speed of the temporal modulation (Fig. 4). An advantage over a modulation-detection threshold paradigm is that stimuli can be presented at supra-threshold levels, allowing for a relatively easy task, suitable for clinical assessments, and for studies with children, or with experimental animals. This is supported by the fact that we did not observe procedural learning effects during the experiments.

Normal Hearing. The observed stMTFs (Fig. 3 and Fig. 4A-E) correspond well with the results of other studies, which usually measured ripple modulation detection thresholds. Chi et al. (1999) and Zheng et al. (2017) measured the full stMTF for normal-hearing listeners using an adaptive modulation-detection threshold paradigm and found band-pass functions for both the spectral and temporal dimensions. They found best ripple detection-thresholds at spectral modulations below or at 1 cycle/octave, and temporal modulations around 4-16 Hz. Despite these small quantitative differences between studies, the general patterns were similar, and in line with our results (Fig. 6): when comparing our promptness data (Fig. 6C,D) with the modulation thresholds collected in the earlier studies (Chi et al. 1999; Zheng et al. 2017 in Fig. 6A,B), the stMTFs resemble each other, at least qualitatively. Note that the color scales are reversed in order, as best responses correspond to high promptness values, but low modulation indices or thresholds. These findings suggest that reaction times are indeed a solid and quick objective measure to determine the spectrotemporal sensitivity of (naïve) listeners.

Chi et al. (1999) proposed a computational model to explain their data, in which the spectrotemporal modulation sensitivity is based on cortical responses to the ripple's spectrogram. The modulation transfer functions generated by their model closely resembled their data, and thus will resemble our data as well. They concluded that 'the upper limits of the spectral and temporal modulation rates are related through the effective bandwidths of the cochlear filters' (Chi et al. 1999).

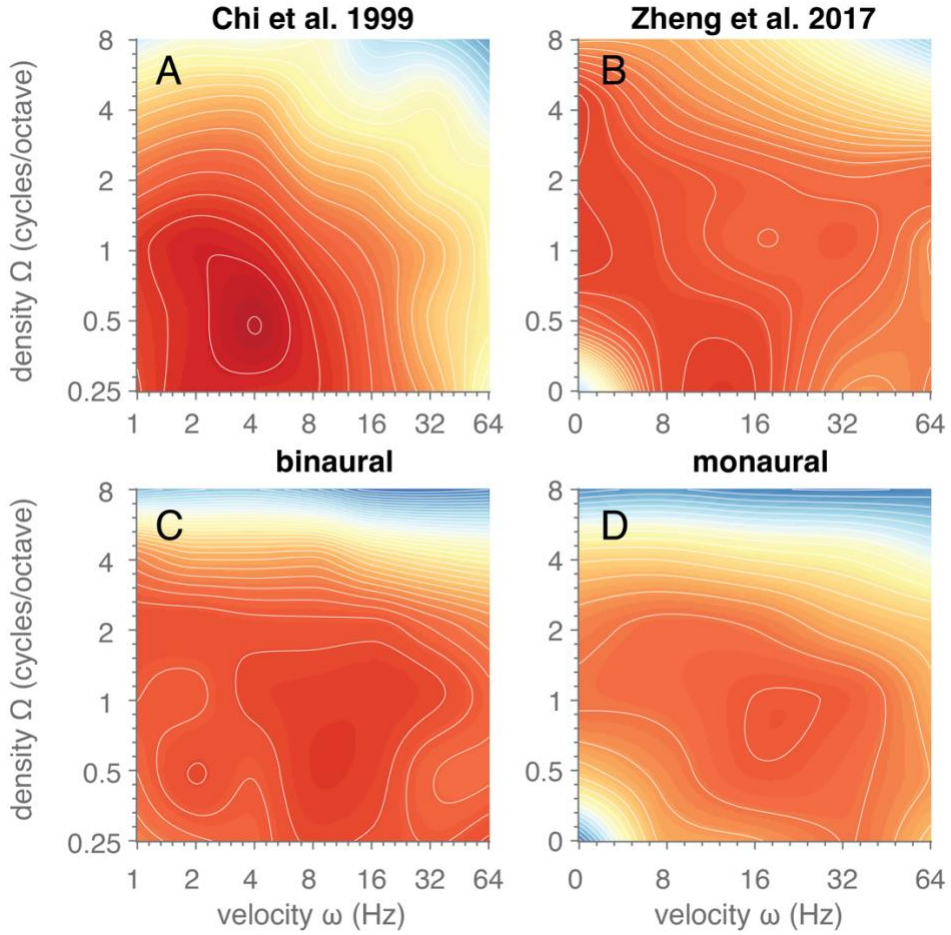

Fig. 6. Comparison of spectrotemporal sensitivity. Images depict stMTFs obtained using **A,B)** a threshold-searching paradigm from the studies of **A)** Chi et al. 1999 and B) Zheng et al. 2017 or **C,D)** a reaction-time task from this study in the **C)** binaural and **D)** monaural listening conditions. Data from previous studies was obtained from Figs. 3 of the respective papers. Image format is the same as in Fig. 4. Data (modulation index) in **A)** was log-transformed to match the data (modulation threshold in dB) in **B)**. Data from this study in C and D was replotted from Fig. 4, matching the velocities and densities used in the previous studies (A vs C, B vs D).

Impaired-hearing simulations. To evaluate our reaction-time test under more challenging listening situations, we manipulated the ripple stimuli using hearing-aid and cochlear-implant simulations. Both simulations made it substantially harder to detect ripple modulations and even impossible for certain parameters, eliciting much longer reaction times compared to monaural normal hearing for all ripples. Bimodal hearing exhibited a 'best-of-both-worlds' effect, following the fastest unimodal condition, which was the cochlear-implant for spectral modulations below 0.75 cycles/octave, and the hearing aid for higher spectral modulations (Fig. 4). An improvement in spectral ripple discrimination for bimodal hearing over the cochlear-implant alone has also been found in users with combined electro-acoustic stimulation in the same ear (Golub et al. 2012).

Other studies have shown a 5-10 dB reduction in the temporal-modulation detection threshold for cochlear-implant users compared to normal hearing (Bacon and Viemeister 1985; Golub et al. 2012; Won et al. 2011), whereas we found a decrease in promptness of $0.4 \pm 0.2\ s^{-1}$ for the well-detectable rates below 16 Hz. In these studies, hearing-impaired listeners performed in between normal hearing and cochlear-implant users. Our hearing-aid simulation, however, showed longer reaction times for temporal modulations (at 0 cycles/octave) compared to the cochlear implant. It should be emphasized that our hearing-aid condition was based on a worst-case-scenario, simulating very little residual hearing, whereas hearing thresholds of the hearing-impaired listeners in the study of Bacon and Viemeister (1985) still reached up to 10-20 dB HL at 1 kHz. Their study also showed a link between degraded temporal sensitivity and reduced listening bandwidth.

Impaired spectral modulation sensitivity with a cochlear implant is a likely result of its band-pass filtering mechanism that limits the spectral information to a set number of spectral bands. Henry et al. (2005) and Berenstein et al. (2008) both found lower spectral ripple modulation thresholds for cochlear-implant users compared to normal-hearing listeners, roughly corresponding to the increased reaction times in our study. Spectral modulation thresholds in hearing-impaired listeners have been reported to be 5-10 dB worse than for normal hearing (Davies-Venn et al. 2015; Summers and Leek 1994), which may agree with the longer reaction times of our hearing-aid simulation compared to normal hearing. A few studies investigated combined spectrotemporal modulation detection thresholds in hearing-impaired listeners, which were often worse compared to normal hearing listeners, especially for low temporal modulation rates (Bernstein et al. 2013; Mehraei et al. 2014; Zheng et al. 2017; Zhou et al. 2020).

Race model. To get insights in the mechanism of combining input at both ears, we used race models to test whether monaural responses could predict binaural performance. For normal-hearing conditions, faster reaction times were elicited when stimuli were presented binaurally compared to monaural presentation. Binaural responses seemed as fast as the race model of statistical facilitation (Fig. 5C). This suggests that ripple detection was determined by a parallel race between the two ears, rather than from neural integration.

As normal-hearing responses were near ceiling performance (fastest possible reaction times, as stimuli were readily detectable), and given that in our simulations there was little spectral overlap between the hearing-aid and cochlear-implant sides, a lack of positive effects in our data on bilateral or bimodal integration, by exceeding race-model performance, may perhaps not be too surprising.

It may be interesting to compare these findings to audiovisual gaze-orienting experiments that aim to study neural integration of visual and auditory signals. Strongest benefits of multisensory interactions (i.e., increased speed, accuracy, and precision of responses) are obtained for stimuli that overlap both in space and time, and thus provide multisensory evidence for a single object. Moreover, these interactions are strongest when the uni-sensory evoked responses are variable and slow (i.e., away from ceiling performance). This phenomenon is known as the 'principle of inverse effectiveness' (e.g. Corneil et al. 2002; van de Rijt et al. 2019; Stein and Meredith 1993; Van Wanrooij et al. 2009). We here propose that beneficial effects of bimodal (cochlear-implant-hearing-aid) integration will depend on whether the auditory system has sufficient evidence that left vs. right acoustic inputs arose from the same auditory object, rather than from unrelated sounds. The strongest bimodal benefits (i.e., enhanced sensitivity) will thus be found: (i) when spectral ranges of cochlear implant and hearing-aid overlap sufficiently (for within-spectral comparisons), and (ii) when monaural reaction-time distributions have sufficient variability and overlap considerably.

In contrast to our results, several studies have shown reaction times to stimuli that exceeded the predictions based on statistical facilitation. However, these studies typically involved responses to multisensory stimuli, or to the dichotic presentation of two spectrally distinct sounds (Gielen et al. 1983; Miller 1982; Schröter et al. 2007; Townsend and Nozawa 1995). Like the findings of Schröter et al. (2007) for auditory stimuli that fused into a single percept, we did not obtain faster responses than expected from statistical facilitation in the bimodal conditions. The benefits of an integrative process must be found outside this study, and likely include the ability to localize sounds and enhancement of speech perception in noisy environments.

Separability. Measuring the stMTF is typically a time-consuming process, for which it would be valuable to know whether the two-dimensional function is simply the product of a temporal and spectral component. A large degree of separability of spectrotemporal sensitivity has been found for normal-hearing listeners and cochlear-implant users (Chi et al. 1999; Zheng et al. 2017). Our data support this notion of large separability (but not 'full' separability) for the normal-hearing and cochlear-implant listening conditions. Nevertheless, the contribution of the second singular value was still substantial, close to 10% in the normal-hearing condition. For the hearing-aid and bimodal listening conditions, the contribution could be as large as 20%, implying a large degree of inseparability. Studies by Bernstein et al. (2013) and Zheng et al. (2017) found that spectrotemporal cues may enhance speech intelligibility and modulation sensitivity in CI users over spectral or temporal modulations alone. These findings hint at some form of spectrotemporal integration, but interestingly this may not be captured by the separability of the stMTF (Zheng et al. 2017).

As can be appreciated from the observable differences between the single and two-component reconstructions (Fig. 4), the second component adds a diagonal interaction component to the two-dimensional MTF. Most of this interaction yields changes to the MTF at 0-density modulations: reactions become slower for low-velocity modulations, and faster for higher velocities. Interestingly, the normal-hearing one-dimensional temporal MTF derived from the 2nd component is nearly a linearly

increasing function of velocity (not shown). This suggests that the 2nd component reflects an amplitude modulation detection rate that can operate faster, simply because the stimulus modulation period is shorter (in line with the linear dependence of monkey reaction times on the amplitude-modulation period, found by Massoudi et al. 2014), and that is independent of how well a modulation frequency can be detected by the auditory system as reflected by the first-component MTF. Taken together, we conclude that a full assessment of spectrotemporal performance requires testing the complete spectrotemporal field of all ripples, by including the second SVD component (Fig. 4).

Conclusion. Reaction times are a solid objective measure for ripple sensitivity. The joint spectrotemporal transfer function closely resembled data from earlier studies that used modulation detection thresholds. Responses to spectrotemporal modulated ripples could be reconstructed by using the first two components of singular value decomposition, suggesting significant spectrotemporal inseparability, especially for the hearing-aid and bimodal listening conditions. We further found that binaural and simulated bimodal reaction times could be predicted from statistical facilitation induced by a race of independent monaural inputs.

Acknowledgements. This work was supported by Advanced Bionics (LCEV), the European Union Horizon-2020 program, ERC Advanced Grant 2016 (ORIENT, nr. 693400; AJVO), and the Radboud University (MMvW). We thank Ruurd Lof, Stijn Martens, and Günter Windau for their technical support and Daisy Louvet for her assistance in data collection.

References

Ausili SA, **Backus B**, **Agterberg MJH**, **van Opstal AJ**, **van Wanrooij MM**. Sound Localization in Real-Time Vocoded Cochlear-Implant Simulations With Normal-Hearing Listeners. *Trends Hear* 23, 2019.
Bacon SP, **Viemeister NF**. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. [Online]. *Audiology* 24: 117–34, 1985http://www.ncbi.nlm.nih.gov/pubmed/3994589 [21 Apr. 2016].
Baer T, **Moore BCJ**. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *J Acoust Soc Am* 95: 2277, 1994.
Berenstein CK, **Mens LHM**, **Mulder JJS**, **Vanpoucke FJ**. Current steering and current focusing in cochlear implants: Comparison of monopolar, tripolar, and virtual channel electrode configurations. *Ear Hear* 29: 250–260, 2008.
Bernstein JGW, **Mehraei G**, **Shamma S**, **Gallun FJ**, **Theodoroff SM**, **Leek MR**. Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *J Am Acad Audiol* 24: 293–306, 2013.
Bernstein LR, **Green DM**. Detection of simple and complex changes of spectral shape. *J Acoust Soc Am* 82: 1587, 1998.
Carpenter RH, **Williams ML**. Neural computation of log likelihood in control of saccadic eye movements. *Nature* 377: 59–62, 1995.
Carpenter RHS, **Reddi BAJ**, **Anderson AJ**. A simple two-stage model predicts response time distributions. *J Physiol* 587: 4051–4062, 2009.
Chi T, **Gao Y**, **Guyton MC**, **Ru P**, **Shamma S**. Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106: 2719–2732, 1999.
Corneil BD, **Van Wanrooij M**, **Munoz DP**, **Van Opstal a J**. Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J Neurophysiol* 88: 438–54, 2002.
Davies-Venn E, **Nelson P**, **Souza P**. Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearinga). *J Acoust Soc Am* 138: 492, 2015.
Elliott TM, **Theunissen FE**. The Modulation Transfer Function for Speech Intelligibility. *PLoS Comput Biol* 5: e1000302, 2009.
Gielen SCAM, **Schmidt RA**, **Van Den Heuvel PJM**. On the nature of intersensory facilitation of reaction time. *Percept Psychophys 1983 342* 34: 161–168, 1983.
Glasberg BR, **Moore BCJ**. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J Acoust Soc Am* 79: 1020, 1986.
Golub JS, **Won JH**, **Drennan WR**, **Worman TD**, **Rubinstein JT**. Spectral and temporal measures in hybrid cochlear implant users: On the mechanism of electroacoustic hearing benefits. *Otol Neurotol* 33: 147–153, 2012.
Henry BA, **Turner CW**. The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. *J Acoust Soc Am* 113: 2861, 2003.
Henry BA, **Turner CW**, **Behrens A**. Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. [Online]. *J Acoust Soc Am* 118: 1111–21, 2005http://www.ncbi.nlm.nih.gov/pubmed/16158665 [21 Apr. 2016].
Hershenson M. Reaction time as a measure of intersensory facilitation. *J Exp Psychol* 63: 289–293, 1962.
Liberman AM (Alvin M. Speech : a special code. 458, 1996.
Litvak LM, **Spahr AJ**, **Saoji AA**, **Fridman GY**. Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J Acoust Soc Am* 122: 982, 2007.
Loizou PC, **Dorman M**, **Tu Z**. On the number of channels needed to understand speech. *J Acoust Soc Am* 106: 2097, 1999.
Massoudi R, **Van Wanrooij MM**, **Versnel H**, **Van Opstal AJ**. Spectrotemporal Response Properties of Core Auditory Cortex Neurons in Awake Monkey. *PLoS One* 10: e0116118, 2015.
Massoudi R, **Van Wanrooij MM**, **Van Wetter SMCI**, **Versnel H**, **Van Opstal AJ**. Stable bottom-up processing during dynamic top-down modulations in monkey auditory cortex. *Eur J Neurosci* n/a-n/a, 2013.
Massoudi R, **Van Wanrooij MM**, **Van Wetter SMCI**, **Versnel H**, **Van Opstal AJ**. Task-related preparatory modulations multiply with acoustic processing in monkey auditory cortex. *Eur J Neurosci* n/a-n/a, 2014.
McDermott JH, **Simoncelli EP**. Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis. *Neuron* 71: 926–940, 2011.
Mehraei G, **Gallun FJ**, **Leek MR**, **Bernstein JGW**. Spectrotemporal modulation sensitivity for hearing-impaired listeners:

Dependence on carrier center frequency and the relationship to speech intelligibility. *J Acoust Soc Am* 136: 301, 2014.
Miller J. Divided attention: evidence for coactivation with redundant signals. *Cogn Psychol* 14: 247–279, 1982.
Moore BC, **Glasberg BR**. Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. [Online]. *J Acoust Soc Am* 110: 1067–73, 2001http://www.ncbi.nlm.nih.gov/pubmed/11519575 [21 Apr. 2016].
Raab DH. statistical facilitation of simple reaction times. *Trans N Y Acad Sci* 24: 574–590, 1962.
Ratcliff R, **Van Dongen HPA**. Diffusion model for one-choice reaction-time tasks and the cognitive effects of sleep deprivation. *Proc Natl Acad Sci U S A* 108: 11285–90, 2011.
van de Rijt LPH, **Roye A**, **Mylanus EAM**, **van Opstal AJ**, **van Wanrooij MM**. The Principle of Inverse Effectiveness in Audiovisual Speech Perception. *Front Hum Neurosci* 13: 335, 2019.
Schröter H, **Ulrich R**, **Miller J**. Effects of redundant auditory stimuli on reaction time. *Psychon Bull Rev* 14: 39–44, 2007.
Shannon R V, **Zeng FG**, **Kamath V**, **Wygonski J**, **Ekelid M**. Speech recognition with primarily temporal cues. [Online]. *Science* 270: 303–4, 1995http://www.ncbi.nlm.nih.gov/pubmed/7569981 [8 Jan. 2014].
Sharma S, **Mens LHM**, **Snik AFM**, **van Opstal AJ**, **van Wanrooij MM**. An Individual With Hearing Preservation and Bimodal Hearing Using a Cochlear Implant and Hearing Aids Has Perturbed Sound Localization but Preserved Speech Perception. *Front Neurol* 10: 637, 2019.
Sharma S, **Nogueira W**, **Opstal AJ van**, **Chalupper J**, **Mens LHM**, **Wanrooij MM van**. Amount of Frequency Compression in Bimodal Cochlear Implant Users Is a Poor Predictor for Audibility and Spatial Hearing. *J Speech, Lang Hear Res* 1–14, 2021.
Stein BE, **Meredith MA**. The Merging of the Sense. Cambridge, MA, US: The MIT Press.: 1993.
Summers V, **Leek MR**. The internal representation of spectral contrast in hearing-impaired listeners. [Online]. *J Acoust Soc Am* 95: 3518–28, 1994http://www.ncbi.nlm.nih.gov/pubmed/8046143 [21 Apr. 2016].
Supin AY, **Popov V V.**, **Milekhina ON**, **Tarakanov MB**. Frequency resolving power measured by rippled noise. *Hear Res* 78: 31–40, 1994.
Townsend JT, **Nozawa G**. Spatio-temporal Properties of Elementary Perception: An Investigation of Parallel, Serial, and Coactive Theories. *J Math Psychol* 39: 321–359, 1995.
Versnel H, **Zwiers MP**, **van Opstal AJ**. Spectrotemporal response properties of inferior colliculus neurons in alert monkey. *J Neurosci* 29: 9725–39, 2009.
Veugen LCE, **Chalupper J**, **Snik AFM**, **van Opstal AJ**, **Mens LHM**. Frequency-dependent loudness balancing in bimodal cochlear implant users. *Acta Otolaryngol* 136: 775–781, 2016a.
Veugen LCE, **Hendrikse MME**, **van Wanrooij MM**, **Agterberg MJH**, **Chalupper J**, **Mens LHM**, **Snik AFM**, **John van Opstal A**. Horizontal sound localization in cochlear implant users with a contralateral hearing aid. *Hear Res* 336: 72–82, 2016b.
Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. *J Acoust Soc Am* 66: 1364, 1998.
Van Wanrooij MM, **Bell AH**, **Munoz DP**, **Van Opstal AJ**. The effect of spatial-temporal audiovisual disparities on saccades in a complex scene. *Exp brain Res* 198: 425–437, 2009.
Won JH, **Drennan WR**, **Nie K**, **Jameyson EM**, **Rubinstein JT**. Acoustic temporal modulation detection and speech perception in cochlear implant listenersa). *J Acoust Soc Am* 130: 376, 2011.
Zheng Y, **Escabí M**, **Litovsky RY**. Spectro-temporal cues enhance modulation sensitivity in cochlear implant users. *Hear Res* 351: 45–54, 2017.
Zhou N, **Dixon S**, **Zhu Z**, **Dong L**, **Weiner M**. Spectrotemporal Modulation Sensitivity in Cochlear-Implant and Normal-Hearing Listeners: Is the Performance Driven by Temporal or Spectral Modulation Sensitivity? *Trends Hear* 24: 1–11, 2020.