Hearing Research 305 (2013) 113–134

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Review

Syllabic (~2–5 Hz) and fluctuation (~1–10 Hz) ranges in speech and auditory processing

Erik Edwards*, Edward F. Chang*

Department of Neurological Surgery, UC San Francisco, USA

ARTICLE INFO

Article history:
Received 22 December 2012
Received in revised form
22 August 2013
Accepted 28 August 2013
Available online 12 September 2013

ABSTRACT

Given recent interest in syllabic rates (~2–5 Hz) for speech processing, we review the perception of "fluctuation" range (~1–10 Hz) modulations during listening to speech and technical auditory stimuli (AM and FM tones and noises, and ripple sounds). We find evidence that the temporal modulation transfer function (TMTF) of human auditory perception is *not* simply low-pass in nature, but rather exhibits a peak in sensitivity in the syllabic range (~2–5 Hz). We also address human and animal neurophysiological evidence, and argue that this bandpass tuning arises at the thalamocortical level and is more associated with non-primary regions than primary regions of cortex. The bandpass rather than low-pass TMTF has implications for modeling auditory central physiology and speech processing: this implicates temporal contrast rather than simple temporal integration, with contrast enhancement for dynamic stimuli in the fluctuation range.

This article is part of a Special Issue entitled "Communication Sounds and the Brain: New Directions and Perspectives".

1. Introduction

A theme of this special issue is the role of vocalizations as stimuli in auditory neuroscience. Vocalizations can be considered as part of a larger class of communication signals used by other species and man-made devices, which by necessity exhibit *modulations*. As Picinbono (1997) states: "Let us remember that a purely monochromatic signal such as $a\cos(\omega t + \phi)$ cannot transmit any information. For this purpose, a modulation is required, …" Likewise, unmodulated noise cannot transmit any information, so we can expect on *a priori* grounds a link between AM/FM (amplitude/frequency modulation) studies and speech studies (Rosen, 1992). In fact, the same auditory regions involved in speech processing are strongly activated by AM/FM sounds. For example, the non-primary cortical areas most activated for AM/FM processing in the syllabic (~2–5 Hz) range are also implicated in pathways for intelligible speech (Scott et al., 2006; Hall, 2012) (Section 4). Thus, we have chosen as our contribution to "Communication Sounds in the Brain" a new consideration of AM/FM processing with relevance to speech. A premise of this review is that careful study of AM/FM results will lead to insights for speech processing.

Recent reviews (Joris et al., 2004; Malone and Schreiner, 2010) cover well the periodicity pitch range surrounding voice fundamental frequency (F_0, ~50–500 Hz), and there is a well-established speech processing literature on extracting F_0. The roughness range (~25–125 Hz) has also been studied extensively and treated well in recent reviews. However, the slower ranges of AM/FM (to be termed the 'fluctuation' range, ~1–10 Hz) are traditionally understudied. We will also find that the neural systems most strongly implicated in fluctuation perception – the 'belt' and 'parabelt' regions of the CNS – are far less studied than 'core' regions (as commented by Goldstein and Knight, 1980; Hall, 2005). In parallel, the slower aspects of speech – syllabic time scales, prosody, stress, intonation, emotional aspects, etc. – are understudied relative to the spectrotemporally-detailed aspects for phonetic purposes. In further parallel, algorithmic approaches to speech processing have only rarely (and more recently) focused on longer time scales. Given the recent interest in syllabic time scales (~2–5 Hz) for speech perception, human neurophysiology, and computer speech processing (Hall, 2005; Greenberg, 2006; Ghitza and Greenberg, 2009; Giraud and Poeppel, 2012; Obleser et al., 2012; Peelle and Davis, 2012), we have chosen to review these time scales in more basic studies of auditory perception and physiology. This is not a comprehensive review of AM/FM sounds, rather a focus on the fluctuation (~1–10 Hz) range and the corresponding time scales of speech. Before embarking on our review, we offer our thoughts on the theme for this special issue.

* Corresponding authors. 675 Nelson Rising Lane, San Francisco, CA 94158, USA.
E-mail addresses: erik.edwards4@gmail.com (E. Edwards), changed@neurosurg.ucsf.edu (E.F. Chang).

0378-5955/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2013.08.017

Abbreviations

AAF	anterior auditory field
AI	primary auditory (cortical field)
AM	amplitude modulation
ASR	automatic speech recognition
BMF	best modulation frequency
CM	caudomedial (cortical field)
CN	cochlear nucleus
CNS	central nervous system
ECoG	electrocorticography
EEG	electroencephalography
FM	frequency modulation
fMRI	functional magnetic resonance imaging
HG	Heschl's gyrus
IC	inferior colliculus
LTI	linear time-invariant
MEG	magnetoencephalography
MGB	medial geniculate body
MGBd	MGB, dorsal division
MGBm	MGB, medial division
MGBv	MGB, ventral division
PET	positron emission tomography
RC	resistance-capacitance
SAM	sinusoidally amplitude-modulated
SFM	sinusoidally frequency-modulated
SNR	signal-to-noise ratio
STMTF	spectro-temporal modulation transfer function
TMTF	temporal modulation transfer function
rTMTF	rate TMTF
vTMTF	vector TMTF
2IFC	2-interval forced-choice

1.1. What are the roles of speech and modulated sounds in auditory neuroscience?

In the exploratory phase of empirical data gathering, speech is a useful stimulus because, amongst other things, it elicits robust activations throughout the auditory nervous system. These yield overall observations concerning directly the stimulus set of interest, which our eventual models must explain. However, the empirical observations available to us – a variety of auditory stations in various species under various anesthetics, using various particular synthetic or natural speech sounds for a given study – do not allow us to easily perceive the essential patterns to be included in the model building exercise. Even a complete catalog of each auditory station responding to each possible phoneme or speech sound would likely remain inadequate. On the other hand, technical stimuli (AM/FM) can be arrayed systematically according to a single parameter (modulation frequency) and related directly to communication theory and signals/systems theory. This obviously accelerates the model building exercise during the difficult early phases, when even the overall layout and essential features of the models are still in question. However, we find that speech is an essential stimulus again in the final stages of model building – the final selection of model structure and specification of model parameters. Since speech is taken to be the stimulus set of interest, the final least-squares or other fit should be determined by the use of speech stimuli whenever possible. We note in this context that speech is usually 'sufficiently exciting', which is a mathematical requirement in system identification (Ljung, 1999), and essentially means that speech is sufficiently rich in spectrotemporal features to cover the signal space of interest. In some contexts, where the modeler has already chosen a certain model structure – for example, the spectro-temporal receptive field (STRF) – then one can usefully skip straight to the use of speech as a stimulus for final least-squares-fit of model parameters. But as we seek more realistic models of CNS function, with new aspects to exploit for speech processing applications, we may require ongoing use of technical stimuli.

We are at a point in history where we have good models of the auditory periphery for most speech processing purposes. By a 'good' model is meant one with broad explanatory power, accurate predictions for arbitrary inputs, and as few parameters as possible (the principle of parsimony). Adequate models of the cochlear nucleus appear to be arriving or nearly on the horizon, but this still places us some distance from a complete computational model of the auditory CNS. Before arriving at a full physiological model, we can hope to arrive at simpler models which are considerably abstracted from actual physiological details (yet including as much physiological insight as possible). In order to approach the modeling problem for auditory CNS, certain simplifications are useful or necessary at the early stages. First, we can ignore binaural/spatial aspects in a first model for speech processing purposes (other than the multispeaker situation, where binaural cues are essential, Cherry, 1953; Bregman, 1990; Schimmel et al., 2008). However, more severe simplifications appear to be required in order to relate psychophysics, human neuroscience, animal neurophysiology, and computer speech processing together in a comprehensible way.

We suggest that studies of modulated (AM/FM) sounds may serve as an intermediate stage, before final model specifications, in the long-term goals of speech neurophysiology and modeling. Extensive bodies of work are already available concerning AM/FM sounds in communication theory, signals and systems theory, human psychophysics, and animal neurophysiology. For a given modulation type, there is a systematic space of signals controlled by a single parameter (modulation frequency), allowing unambiguous mapping across research domains into a single orderly framework. While this is still not a sufficiently complicated space to understand all aspects of speech, it is a long way in the right direction compared to clicks and tones, the traditional technical stimuli. The fact that many workers have adopted modulation filter banks (Kay and Matthews, 1972; Dau et al., 1997) or related approaches (Greenberg and Kingsbury, 1997), i.e. adding to the spectral and temporal dimensions a modulation dimension (Atlas and Shamma, 2003; Singh and Theunissen, 2003), indicates the utility of having a stimulus set which can be systematically ordered along the modulation frequency axis (as opposed to various random stimuli).

AM/FM sounds also have the advantage of lacking spectral structure. We noted that severe simplification is often required at early model building stages, such as ignoring binaural/spatial processing. We can also ignore spectral-domain pitch processing for AM/FM sounds below the range where periodicity pitch is elicited (below ~50 Hz), where the resulting spectral structure is not resolvable by the ear. Thus, we can ignore two-tone interaction, lateral inhibition, and other complexities of cross-spectral processing. Results discussed below (Sections 2 and 4) indicate that spectro-temporal processing is to a first approximation *separable*, such that spectral and temporal processing studied separately can be recombined to predict spectro-temporal results. Before auditory CNS models will become available for arbitrary signals, preliminary models to account for temporal stimuli are likely to appear. Since speech can be understood by temporal cues alone (Shannon et al., 1995), this further suggests that study of temporal processing in

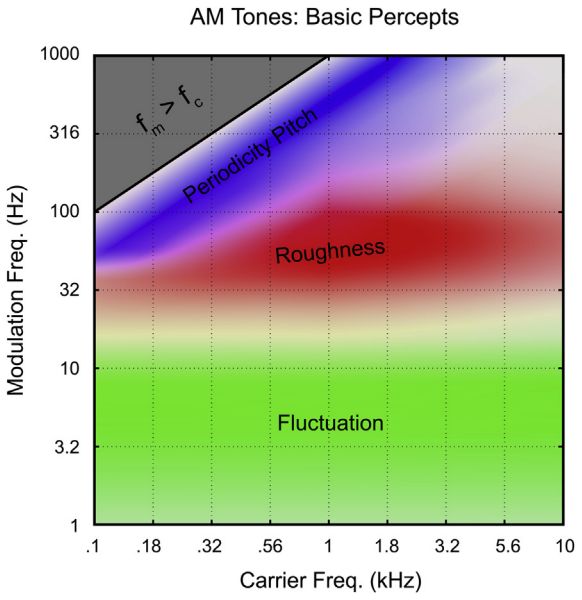

Fig. 1. The three basic perceptual qualities experienced during listening to AM tones, based quantitatively on the magnitude scaling data of Fastl (1983) and Fastl and Zwicker (2007). All results were obtained at a comfortable listening level (e.g., 70 phon), with 100% AM depth. Note that the periodicity pitch and roughness ranges overlap (purple color). The peak of pitch strength is always experienced when $f_m = f_c/2$, because in this case the lower sideband is positioned at precisely the fundamental frequency. Roughness is defined relative to a standard of $f_c = 1$ kHz and $f_m = 70$ Hz. Similar overall results are obtained for AM noise and FM tones, except that pitch strength is much weaker for AM noise, and roughness is much stronger for FM tones. Note that the identical percepts are elicited by speech stimuli in so far as the spectrogram exhibits modulations in the appropriate ranges. For speech, 'periodicity pitch' is 'voice fundamental', and 'fluctuation' is sometimes called 'rhythm'.

isolation from complex spectral structure may serve as a first approximation for preliminary models. However, as we argued above, these models should then be tested for parameter specification by use of natural speech signals when possible.

Finally, there is abundant evidence that the same basic auditory percepts experienced during listening to modulated sounds ('fluctuation', 'roughness', 'periodicity pitch') are also experienced when the same modulation frequencies are present in the speech signal. Voice fundamental frequency (F_0), from glottal pulse rate, elicits the same basic pitch sensation as periodic clicks or AM/FM stimuli of the same frequency. Glottal shimmer (AM) and jitter (FM) result in roughness range (~25–125 Hz) modulations, and correspondingly elicit a perception of roughness in the voice (Wendhal, 1966a,b; Coleman, 1971). However, tremulo (AM) and vibrato (FM) in the voice occur below the roughness range (~2–20 Hz, usually ~7 Hz), and generally sound pleasing and form part of musical technique (Seashore, 1936; Potter et al., 1947). Thus, perception of AM/FM sounds directly predicts perception of vocalizations, in so far as the same modulation rates are present. In Section 2 we consider these basic percepts for AM/FM sounds, where it should be kept in mind that these are the same basic percepts experienced during listening to vocalization stimuli.

2. Basic auditory percepts for AM and FM sounds

Before narrowing our focus to the fluctuation range (~1–10 Hz), we set up the context of the full range of AM/FM percepts.

2.1. AM tones

The most basic modulated sounds are "beats", which consist of two sinusoidal tones at frequencies f_1 and f_2 added together. The resulting sound exhibits AMs at the frequency $f_m = |f_1 - f_2|$. Beats

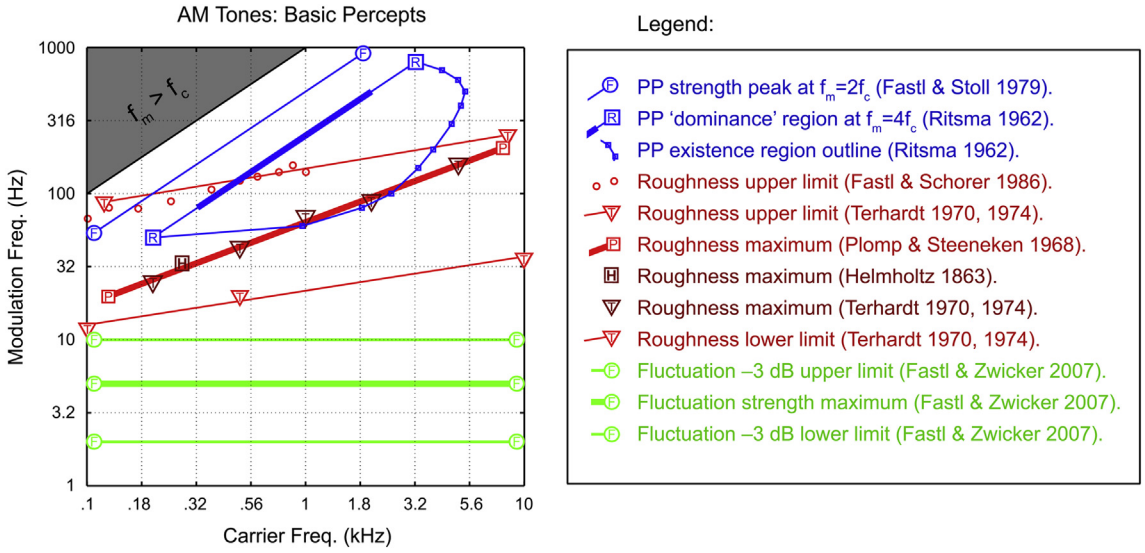

Fig. 2. The three basic percepts during listening to AM tones, according to various authors. The thin lines can be taken to indicate boundaries of *existence regions* and the thick lines can be taken to indicate a maximal or dominant region for the percept. *Fluctuation (green)*: The thick green line indicates the peak of fluctuation strength at 5 Hz, and thin lines indicate approximately the −3 dB points. *Roughness (red)*: The upper/lower limits of roughness (thin red lines) are obtained from Terhardt (1970, 1974). The small red circles are from a re-examination of the upper limit of roughness by Fastl and Schorer (1986). The data of Plomp and Steeneken (1968) on maximal roughness were well-fit by a straight line (in log–log coordinates), given by the thick red line. The dark-red triangles give the maximal roughness according to Terhardt (1974), in close agreement. Interestingly, the original point of maximum roughness (German: 'Rauhigkeit') as given by Helmholtz (1863) for two beating violin tones (dark-red square), is also in close agreement with Plomp and Steeneken (1968). *Periodicity Pitch (blue)*: The thick blue line is the classic 'dominance region' of Ritsma (1962), drawn according to his rule that the periodicity pitch percept is dominated by frequencies near the 4th harmonic, $f_c = 4f_m$. This also matches the rule for pitch dominance with ripple noise (at $4/\tau$) (Bilsen and Ritsma, 1967; Yost et al., 1978; Yost, 1982). The small blue circles and blue arc delineate the existence region of periodicity pitch from Ritsma (1962), each point obtained as the average over his 3 subjects. Along this line, a 100% modulated tone just evokes a sensation of periodicity pitch (note that he did not include $f_m = f_c/2$ on conceptual grounds, arguing that it did not qualify as a 'residue pitch' given that the fundamental frequency is present). The upper blue line (Fastl and Stoll, 1979; Fastl and Zwicker, 2007) is drawn at $f_m = f_c/2$.

were already understood by the time of Helmholtz (1863), see Wever, 1929 for history, who introduced the term 'roughness' for two violin notes beating at $f_m = \sim 30$ Hz. A second means of generating AM tones is to multiply a pure tone by an envelope of frequency f_m. The perception of these two types of AM sound is essentially identical and they are summarized together (Figs. 1 and 2). To our knowledge, only one author has used magnitude scaling, over the full range of AM (from fluctuation through roughness to periodicity pitch), and thus obtained a self-consistent and fairly comprehensive data set (Fastl, 1977, 1982, 1983; Fastl and Stoll, 1979; Fastl and Zwicker, 2007). In Fig. 1, we summarize Fastl's data for AM tones, as obtained carefully from figures in his text (Fastl and Zwicker, 2007). Fig. 2 includes a variety of relevant data collected over the decades for comparison, and for delineating existence vs. non-existence regions.

Note that the region between fluctuation and roughness is not adequately covered in Fig. 1 ($f_m = \sim 16$ Hz). The percept around 16 Hz does not seem prototypical of fluctuation or roughness as currently defined; we suggest the use of "intermittence" (Wever, 1929) or "flutter" (Nourski and Brugge, 2011) for this range. Prototypes for the 4-category scheme could be 1-kHz tones modulated at 4, 16, 64, and 250 Hz. This would yield logarithmic spacing on the f_m scale; for example, modulation filter banks typically employ logarithmic spacing above the fluctuation range (Dau et al., 1997).

2.2. Other modulated sounds

The preceding section directly concerned AM tones (including beats), but the results apply directly to FM tones and AM noise. FM and AM tones with modulation rates in the periodicity pitch range are perceptually indistinguishable – both exhibit strong sidebands separated by a sufficient degree that they are resolvable by the ear. Thus, the blue regions in Figs. 1 and 2 can be considered identically applicable to AM and FM tones, where a pitch sensation is evoked by the spectral (harmonic) structure.

For the roughness range, FM tones have been studied specifically by Terhardt (1968) and Kemp (1982). Both authors emphasize the close similarity between the roughness results for AM and FM. The major difference is that FM tones elicit a greater roughness percept (up to 6 times greater: Fastl and Zwicker, 2007), but this does not appear to involve any shift in the existence region or region of maximal roughness. Thus, the red regions in Figs. 1 and 2 can be considered equally applicable to AM and FM tones.

For the fluctuation range, FM tones have only been studied quantitatively by Fastl (1983) and Fastl and Zwicker (2007) to our knowledge. His results were obtained for AM and FM in the same subjects and the comparisons again indicate no appreciable difference between AM and FM. Obviously, we can clearly distinguish AM and FM perceptually, but their fluctuation strengths have a similar tuning as a function of f_m, peaking near ~ 5 Hz. Thus, the green regions in Figs. 1 and 2 apply to both AM and FM tones.

For AM broadband noise, the results are similar to those for AM tones within the frequency range from $f_c = \sim 0.4$–4 kHz (wherein the most detailed auditory processing occurs, and which is most critical for speech intelligibility). That is, the frequencies near ~ 1 kHz are most strongly weighted in determining the outcome for broadband noise. Thus, broadband noise with AM near ~ 4 Hz gives a strong fluctuation percept (Fastl, 1982), and AM noise near 70 Hz (or ripple noise with delays near $\tau = 1/70$ s) gives a strong roughness percept (Patterson et al., 1978; Bilsen and Wieman, 1980). The major difference between AM tones and AM noise is a strong reduction in the periodicity pitch strength for AM noise. Although melodies can be recognized using only AM noise (Burns and Viemeister, 1976, 1981), where the long term spectrum remains white, only a faint, "whispy" pitch-like sensation is evoked.

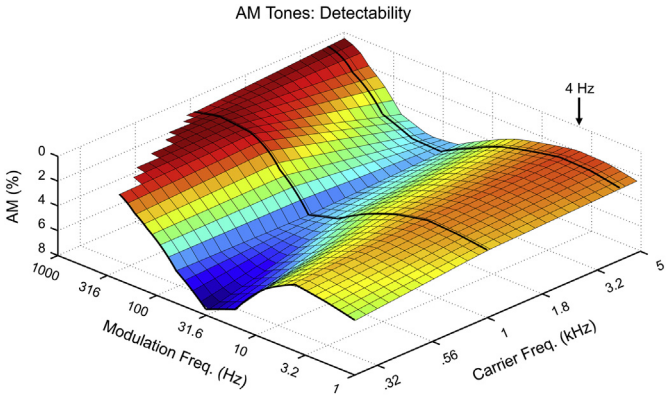

Fig. 3. Detectability of AM tones as a function of f_c and f_m according to data of Zwicker (1952), which is the most complete to date. Three f_cs were tested (0.25, 1 and 4 kHz), as indicated by the black curves, over a wide range of f_ms. Further details are given in the text. The major result for present purposes is the peak in sensitivity centered at $f_m = 4$ Hz. Note also the overall increase in sensitivity in going to toward the basal region of the cochlea (higher f_cs), which play an overall stronger role in temporal envelope processing.

However, this does not appear to involve any overall shift of the existence or dominance regions. For example, we noted in Fig. 2 that the dominance region of Ritsma (1962) for AM tones is confirmed with cosine noise (Bilsen and Ritsma, 1967; Yost et al., 1978; Yost, 1982).

Overall, the basic percepts illustrated in Figs. 1 and 2 are similar for all technical stimuli – beats, AM tones, FM tones, AM noise, and ripple noise – thus increasing their utility as summaries of many basic psychoacoustic results. As already mentioned, speech stimuli also elicit the same basic percepts in so far as they exhibit modulations in the ranges indicated in Figs. 1 and 2.

3. Focus on the fluctuation range

3.1. AM detectability

Having established the overall percepts for AM and FM sounds, we focus in this section on the fluctuation range ($\sim$1–10 Hz). Specifically, we will survey a body of evidence in support of the central claim of this review – that human auditory perception exhibits a tuning to modulations occurring within the fluctuation range, peaking broadly at $\sim$2–5 Hz. This is similar to the typical syllabic rate of speech, to be discussed in Section 3.3.

To preview this claim, we make another plot in the form of Figs. 1 and 2, except now instead of depicting the *strengths* of the percepts (Fig. 1) or the approximate *existence regions* of the percepts (Fig. 2), we plot the *detectability* of AM as a function of carrier frequency (f_c) and modulation frequency (f_m). While this does not allow insight into the perceptual experience of the listener, it has the advantage of requiring no verbal labels or subjective categories. Instead, the listener merely needs to indicate whether modulation has or has not been detected in some sound; for a pair of sounds, the listener indicates which one was modulated vs. unmodulated. To date, the most comprehensive data of this type remains that of Zwicker (1952), who studied four subjects at several different loudness levels (Ls) and f_cs. For each combination of L and f_c, a wide range of f_ms were tested from 1 Hz to several kHz. Fig. 3 shows the outcome for $L = 60$ phon (a comfortable listening level in the range of conversational speech); note in the plot that a peak represents *maximal detectability*. Notice that there is a broad trough (difficult to detect AM) in the middle of the roughness range, rising on either side to two broad peaks. The first peak occurs in the periodicity pitch range, and depends on spectral processing (i.e., the sidebands become detectable as separate pitches or as part of a harmonic

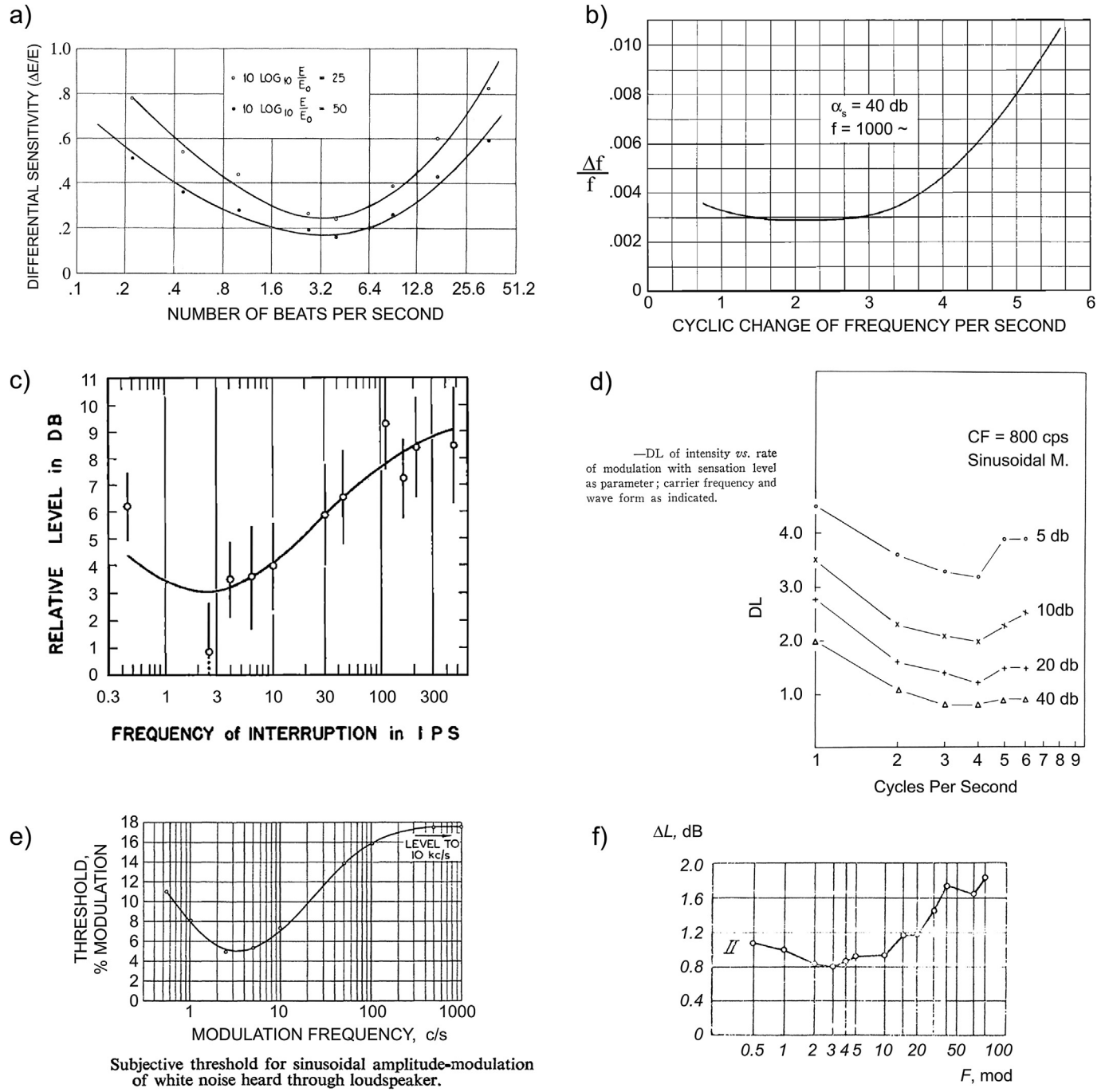

Fig. 4. Historical demonstrations of maximum sensitivity to modulations in the range ~2–5 Hz. See text for info. a) Riesz (1928), beats (SAM tones). b) Shower and Biddulph (1931), FM tones. c) Pollack (1951), interrupted white noise. d) Tonndorf et al. (1955), SAM tones. e) Stott and Axon (1955), SAM white noise. f) Dubrovskii and Tumarkina (1967), SAM white noise.

pattern), but this is beyond the present scope. The second peak occurs at 4 Hz for each carrier frequency tested (tests were at $f_m = 1, 2, 4, 8, \ldots$ Hz). Importantly, the sensitivity declines at 2 Hz and further at 1 Hz. Had Zwicker tested lower, he likely would have found the detectability of AM to decline even more drastically, as seen below (Section 3.2).

Zwicker's finding at ~4 Hz (also, Zwicker and Feldtkeller, 1967) is not widely appreciated as an established fact about human AM processing. The major reasons are to be discussed in Section 3.3 (emphasis on the roughness and periodicity pitch ranges, and the claims of a low-pass TMTF). Given that evidence in favor of a (broad) band-pass tuning in the range ~2–5 Hz has obvious implications for syllabic rate speech perception (Section 3.3), we next enumerate a comprehensive survey (but kept as succinct as possible) of the psychoacoustic evidence for this central claim.

3.2. The evidence

Claim. Human auditory perception of modulated sounds – AM tones, FM tones, AM noise, and ripple sounds – exhibits a (broad) band-pass tuning of sensitivity in the range ~2–5 Hz, and is not simply a low-pass response.

Evidence

1. In perhaps the first modern psychoacoustic experiment with electronic equipment, Riesz (1928) set out to determine the sensitivity of the ear to small differences in intensity. Rather than creating abrupt increments of intensity, the method of beating tones was used to create smooth intensity modulations. Riesz first tested a range of f_cs and f_ms in 3 observers to find the region of best sensitivity: "All observers showed practically the same results at all frequencies and intensities. A representative curve… is shown in Fig. [4a] (the particular frequency used here was 1000 cycles per second). It is characterized by a broad minimum in the neighborhood of 3 cycles of intensity fluctuation per second." See Fig. 4(a).
2. Shower and Biddulph (1931) set out to determine the sensitivity of the ear to small differences in frequency. Like Riesz (1928), abrupt transitions were avoided by sinusoidal modulation: "Since it is impossible to vary the frequency of a system without scattering energy into frequency regions other than that being used, a method of variation in which this scattering would be a minimum was sought." Different rates of sinusoidal FM were tested to determine the rate of maximum sensitivity to subtle variations of the tone frequency: "The results of these observations are shown in Fig. 4. The curve shows a broad minimum from 2 to 3 variations per second." See Fig. 4(b). We note that tasks requiring acute pitch sensitivity tend to reveal the slower end of the sensitivity range ($\sim$2–3 Hz), so the fact that Riesz's minimum for AM was at $\sim$3–4 Hz may be significant.
3. Pollack (1951) studied interrupted white noise and found "a broad minimum in the region of 4 i.p.s. (interruptions per second)". This is not a standard AM detectability experiment (e.g., it used loudness judgments), but it has been cited for the earliest premonitions of the system's analysis viewpoint (van Zanten, 1980). It is also interesting that Pollack discussed his results in terms of the then-current 'alpha-scanning' hypothesis of brain rhythms, since the equivalent vision experiment with interrupted white light gives a peak near the alpha range (Bartley, 1939). This was also related to growing interest in "excitability cycles" (Clare and Bishop, 1952; Chang, 1960), and thus premonitory of current writings on the role of cortical theta oscillations in auditory/speech processing.
4. Zwicker (1952) tested a wide range of sinusoidal AM (SAM) tones. These were produced by multiplying a tone by a sinusoidal envelope, whereas beats are produced by summing two nearby tones. However, the results are very similar throughout the f_m vs. f_c plane (Section 2), so we refer to both as "AM tones" or "SAM tones". As covered in Section 3.1 (Fig. 3), Zwicker found peak AM sensitivity at 4 Hz.
5. Tonndorf et al. (1955) used SAM tones to test the difference limen for intensity (DL, synonymous for our purposes with the just-noticeable difference). This is very similar to Riesz (1928), but measures were obtained in 19 subjects and focused on the AM range $f_m = 1$–6 Hz. They found (see their Fig. 5 in Fig. 4d): "As seen in Figure 5, the variation with modulation frequency was similar for all sensation levels, reaching its smallest value at 4 cps, although the difference between 3 and 4 cps was rather small. In a similar manner, the between-subject variation reached a minimum at 4 cps, …"
6. Stott and Axon (1955) provided an important expansion of the above results to broadband noise and to FM. They tested 8 subjects with tones from $f_c = 0.05$–10 kHz, for both AM and FM, as well as SAM broadband noise. They made an important methodological comment (Section 3.4) that just presenting sounds and asking the subject if they notice the presence modulation is not an optimal method: "…aural fatigue and auditory imagery were serious factors in these conditions. …Greater consistency resulted if the pure tone was presented first and the modulation gradually increased until the subject indicated that he was aware of the change." For AM tones, they found: "There is enhanced perception of modulation frequencies around 3 or 4c/s, but below 0.5c/s perception becomes more difficult as memory is called into play." But for FM tones, the maximal sensitivity was found around 2–3 Hz, confirming Shower and Biddulph (1931). They were the first to test SAM noise, and found: "As with pure tone, the most sensitive discrimination is found in the region of 3–4c/s, where the threshold is 5%." See Fig. 4(e).
7. Dubrovskii and Tumarkina (1967) studied SAM broadband noise and the subject indicated "the time at which he recognized the presence of modulation of the signal." They found (Fig. 4e) that: "The curves attain a minimum in the range of modulation frequencies 1.5–5 cps." This study is noteworthy for being the first to suggest a model for the low-pass aspect of the curve (the decreasing sensitivity above 5 Hz): "For low modulation frequencies (on the order of 2–5 cps) the ear manages to keep up with the variations in noise level. With an increase in modulation frequency, the variations in the level become too rapid for the stimulus rise and fall processes in the auditory system to be able to keep pace with the level changes. In this case…the difference between the minimum and maximum excitation diminishes." That is, the output of the integration (excitation in the CNS) should exhibit less amplitude modulation than the input signal, for faster AM rates. They show a simple RC integration circuit as a model.
8. Zwicker and Feldtkeller (1967) report extensive measurements with AM and FM tones, AM white-noise, and AM and FM bandpass noise, including the data of Zwicker (1952). They also provide a clear introduction to AM and FM in general, so this is a good starting point for an introductory reader (English translation, Zwicker and Feldtkeller, 1998). They found for AM tones, AM white-noise, and AM bandpass-noise that: "The highest sensitivity is at a modulation frequency of 4 Hz." For FM bandpass noise: "As for tones, the ear is most sensitive to modulation frequencies between 2 and 5 Hz." And for FM tones: "All the curves have a broad minimum in the range of 2 to 5 Hz." We show their results for AM white noise in Fig. 4(e), because these exhibit an important difference compared to the results for AM tones (Fig. 3). Although the region of maximal sensitivity in the fluctuation range is essentially unchanged, the sensitivity to periodicity-pitch range AM is not present for white noise as it is for tones. This is further confirmation that the sensitivity to periodicity-pitch range AM is based primarily on spectral processing, since this cue is not available for AM noise as it is for AM tones.

This concludes the classic evidence for the claim of $\sim$2–5 Hz tuning (noting that this is not a sharp peak, and that frequencies below 1–2 Hz must be tested to clearly see the full bandpass nature). An important summary point is that the same general finding applies to all technical stimuli tested (AM and FM tones and narrow-band noise, and AM broad-band noise). We have omitted a few references of lesser historical value (such as abstracts), but some of these can be found in the review of Kay (1982). Further evidence is found in studies of spectro-temporal modulation transfer functions (Section 3.5), but first we must introduce the temporal modulation transfer function (TMTF).

3.3. TMTF and relevance to speech

An important concept required for further evidence on the $\sim$2–5 Hz tuning, and its relevance to speech, is the temporal

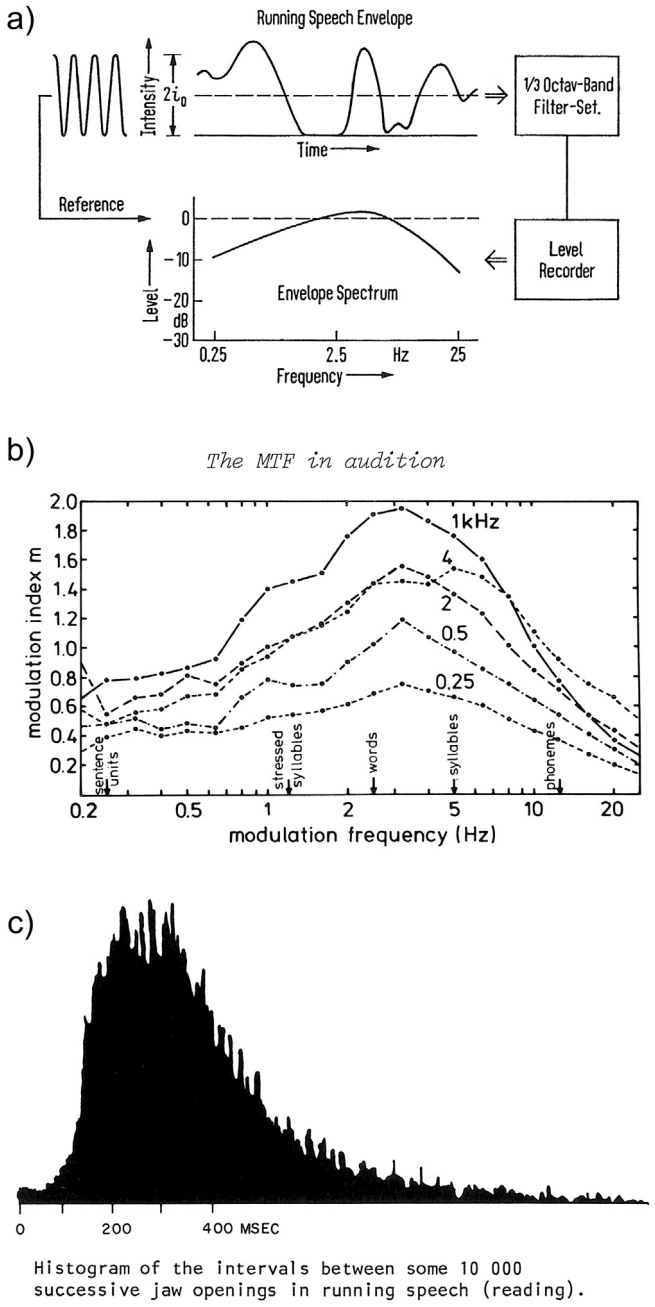

Fig. 5. The envelope spectrum of natural speech production. a) Houtgast and Steeneken (1973): "The fluctuations of running speech as represented by the envelope spectrum." b) Plomp et al. (1984): Average envelope spectrum for 1-min discourses from 10 male speakers. c) Ohala (1975): Jaw opening intervals during continuous speech. The majority of intervals occur in the range ~200–500 ms (i.e., ~2–5 Hz) and almost all intervals (other than small motion noise discussed by the author) occur in the range ~100–1000 ms (~1–10 Hz).

modulation transfer function (TMTF). The TMTF was first introduced into hearing research by Møller (1972a,b), who studied the responses of single-units in the cochlear nucleus to AM and FM stimuli (Section 4). The concept of the TMTF is quite simple: take an input signal and an output signal, related by a system (black box); but instead of relating the *raw* input/output signals, we instead attempt to relate the *envelopes* of the input/output signals. It is that simple – extract the envelopes of the input and output, and compute a transfer function. For Møller's TMTF, the input was the envelope of the stimulus (AM tones or noises) and the output was the time-varying firing-rate of the single-unit (like the envelope, the firing-rate is a non-negative quantity, and so behaves like an envelope for computing a TMTF).

Independently, Houtgast and Steeneken (1973) introduced the TMTF in the context of room acoustics. Typical rooms result in a low-pass smoothing of the envelope of acoustic signals, with important implications for speech processing. For example, this smoothing most strongly reduces AM in the periodicity pitch range (~50–500 Hz), but does not affect the spectral pattern (harmonic structure), so it makes sense that we perceive voice fundamental primarily by spectral rather than temporal processing. As part of this work, Houtgast and Steeneken (1973) and Houtgast et al. (1980) computed the long-term envelope spectrum of speech. That is, they extracted the (overall) intensity envelope of the speech waveform, and computed its spectrum. They found that the modulation spectrum of speech exhibits a broad peak in the range ~2–5 Hz (Fig. 5). In case there is any doubt that this abstract measure of the acoustic envelope represents the syllabic rate of speech production, we include the concrete measurements of Fig. 5(c) (Ohala, 1975) where: "The subject (…) read technical prose for about 1 ½ hours; jaw movement was tracked optically… There is a large single peak around 250 ms, which may be the modal syllable rate or the preferred frequency of the mandible."

In an important continuation of the TMTF work, Drullman et al. (1994a,b) studied the *manipulation* of the speech envelope spectrum in terms of its consequences for speech perception and intelligibility. Specifically, Drullman et al. either low-pass filtered (1994b) or high-pass filtered (1994a) the Hilbert envelope of speech, within each of the sub-bands separately, and then reconstructed the speech using the filtered envelope and the original 'fine structure'. Intelligibility was degraded primarily by removing AM in the fluctuation range (~1–10 Hz, peak in the range ~2–5 Hz), with some additional degradation for consonants at ~8–16 Hz. Although there are certain technical difficulties in using the Hilbert envelope directly in this application (e.g., see Clark and Atlas, 2009), their results have been overall confirmed and are a major historical impetus behind the current interest in the ~delta/theta bands for speech processing. A related historical impetus from the same era was the finding of Shannon et al. (1995) that speech devoid of spectral structure (temporal envelope cues only) can remain intelligible.

3.4. The claim of a low-pass TMTF

Given the extensive body of evidence for ~2–5 Hz modulation sensitivity in psychoacoustical studies and the obvious relevance to speech, it might seem impossible that a psychoacoustician of the 1970s or 1980s would overlook this evidence, and instead see a purely *low-pass* response for AM detectability. Yet the majority of workers appeared to turn to the low-pass view in the late 1970s, and we are still partly in the era of vaguely accepting the existence of a low-pass TMTF. We argue here that, in fact, little or no evidence in favor of a low-pass TMTF was produced. Once the low-pass view became prevalent, and interest began to rise for the 40-Hz AM range, the low-pass view became confirmed for a trivial reason – many studies on AM processing did not use f_ms lower than 5–10 Hz, and so could not possibly have detected the band-pass nature of the tuning centered at ~2–5 Hz. In order to see how the era of low-pass TMTF came about, we must enter briefly into the auditory model which these workers were attempting to confirm.

Licklider (1959) introduced the following basic model of the peripheral auditory system: the acoustic stimulus is subjected to band-pass filtering (according to the cochlea), and then half-wave rectified and smoothed (according to the conversion from hair-cell to auditory-nerve response). This basic model, including a

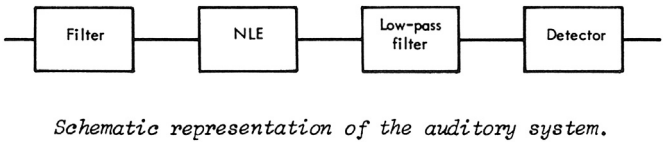

Fig. 6. Schematic model of the auditory system from Rodenburg (1977): "We assume that the auditory system can be described by a model consisting of a critical band filter, a nonlinear element (rectifier), a low pas filter and a detector. The modulation threshold is determined by a low-pass filter and a detector."

pre-emphasis stage (according to the middle ear), was given again by Flanagan (1961). Both Licklider and Flanagan were highly influential in auditory and speech theory, and this basic model has since been used innumerable times, with various choices for the filters. Now, the half-wave rectified and smoothed stimulus is 'the envelope' according to the model auditory system (even though it intermixes 'fine structure' according to Hilbert transform theory), and so the final stage of smoothing should result in a low-pass response of the auditory PNS with respect to AM processing. It is easy to see how this highly influential model leads to the expectation of a low-pass TMTF. If the smoothing time-constant were, say, 10 ms, then modulations occurring within this effective duration, i.e. $f_m > 100$ Hz, would be eliminated or reduced by the smoothing. Another way of stating this is that our *temporal acuity* (Green, 1973) is limited by the smoothing action of the hair-cell/synapse. Green's student, Viemeister, would later become one of the leading authors on auditory temporal processing, still highly cited today. Viemeister's work, along with two other early authors on the TMTF in psychoacoustics (Rodenburg, 1977; van Zanten, 1980), forms the primary historical origin of the notion (still assumed, implicitly or otherwise, by many current authors), that human perception of AM sounds is basically a low-pass process. We now take a closer look at these three early TMTF authors, and show that in fact they produced little or no evidence for a strictly low-pass TMTF in AM processing.

1. The TMTF was first introduced into psychoacoustics by Rodenburg (1972, 1977), who studied the threshold for detecting modulated vs. unmodulated white noise (2 interval forced-choice, 2IFC). Recall from Section 3.2 that there is no AM sensitivity in the periodicity-pitch range for white noise, so the sensitivity for AM noise declines monotonically through the roughness range and into the periodicity-pitch range. Thus, starting at the peak at ~2–5 Hz and higher, we expect a purely *low-pass* appearance for AM white noise. This is exactly what Rodenburg (1977) found, and this is particularly expected given that the great majority of his data was collected in the range $f_m = 5$–1000 Hz. His Figure 2 shows two isolated data points for AM sensitivity in the range 2–4 Hz, and these actually *do* exhibit a decline in sensitivity relative to the 5–10 Hz range. Within the range of variability displayed, the bandpass functions of Section 3.2 would probably fit his data equally well. Since Rodenburg assumed that the AM threshold was determined by the low-pass filter in his model (Fig. 6), he fit a simple RC-filter characteristic to his data (note that an RC filter is a low-pass smoothing filter and also called a 'leaky integrator').
2. The next student of the TMTF in psychoacoustics was Viemeister (1973, 1977, 1979), see also Bacon and Viemeister (1985), Viemeister and Plack (1993). Like Rodenburg, Viemeister (1977) was driven by the Licklider–Flanagan model: "According to this scheme a single frequency channel consists of a bandpass, "critical band" filter followed by a nonlinearity, typically half-wave rectification, followed in turn by a lowpass filter. In the context of this descriptive model the present problem is to measure the transfer function for the lowpass filter…" Like Rodenburg, Viemeister (1977) used SAM white noise in a 2IFC experiment (the subject indicates which interval contains the modulation), and fit the data with an RC filter characteristic. However, the data shows a subtle decline in sensitivity going from $f_m = 4$ Hz to 3 Hz, and again from 3 Hz to 2 Hz. Within the range of variability displayed, a bandpass characteristic would fit the data equally well. Moreover, Viemeister discloses a potentially serious methodological flaw with the 2IFC procedure: the modulator is gated at sine phase, such that at the very onset of the modulated interval the intensity is at its lowest point. This provides a simple onset detection cue which will be stronger the slower the modulation given that subjects are sensitive to stimulus rise time (unfortunately, Rodenburg and van Zanten do not specify the onset phase for their stimuli). Also recall the methods comment of Stott and Axon (1955): the 2IFC procedure is expected to fatigue the subjects and give results with greater variability than the method used in most of the classic studies of Section 3.2 (where the sound was turned on and then the modulation depth varied until just detectable).
3. van Zanten (1980) also used the model of Fig. 6, and assumed that the TMTF was a measure of "the transfer function of the leaky integrator". Van Zanten used similar methods as Rodenburg and Viemeister, and similar to their data, a subtle increase in AM detection threshold is sometimes seen at the lowest AM frequency tested (2 Hz), possibly consistent with the bandpass model, particularly given the variability displayed in the data for 3 subjects. However, he also concluded in favor of a low-pass characteristic.
4. Fastl (1977) presented an extensive study of temporal masking: an AM noise is presented as a *masker*, and the task is to detect a brief probe tone presented within the noise. If the probe tone is presented at the peak of the noise, there is an increased detection threshold compared to if the probe tone is presented in the trough of the noise. Fastl did not obtain measures below an AM of 5 Hz, but it is should be obvious that the detection of the probe can only continue to improve with lower frequency maskers (because the probe tone will be surrounded by longer and longer intervals of near silence). Since the detectability of AM is not tested here (just the detectability of a probe tone as a function of local-SNR), we do *not* consider this a measure of "the TMTF". Nonetheless, both Rodenburg (1977) and Viemeister (1977) used this experiment as a second measure of "the TMTF". As expected, they both found a low-pass function. Examination of their actual data shows a peak at 4–5 Hz AM, and not at 2 Hz, their lowest frequency tested, which is surprisingly compatible with the band-pass view.

Thus, we find that the early students of the TMTF were driven by the Licklider–Flanagan model of the auditory periphery (Fig. 6), and essentially viewed the TMTF as an exercise in finding the integrating time constant of the low-pass filter element (including CNS contributions). These studies were highly influential and initiated what we might term 'the low-pass era' for auditory temporal processing. Other influential authors of that era studied AM detectability only for frequencies above 5 Hz (Terhardt, 1974) or even 20 Hz (Patterson et al., 1978), and the discovery of the 40-Hz EEG response by Galambos et al. (1981) decisively shifted interest away from the lowest modulation frequencies. However, we may be entering a new era, partly instigated by interest in syllabic-rate speech processing (Introduction), where occasional acceptance of the classic bandpass characteristic is indicated. Another sign of a return to the bandpass interpretation comes from recent studies of the STMTF.

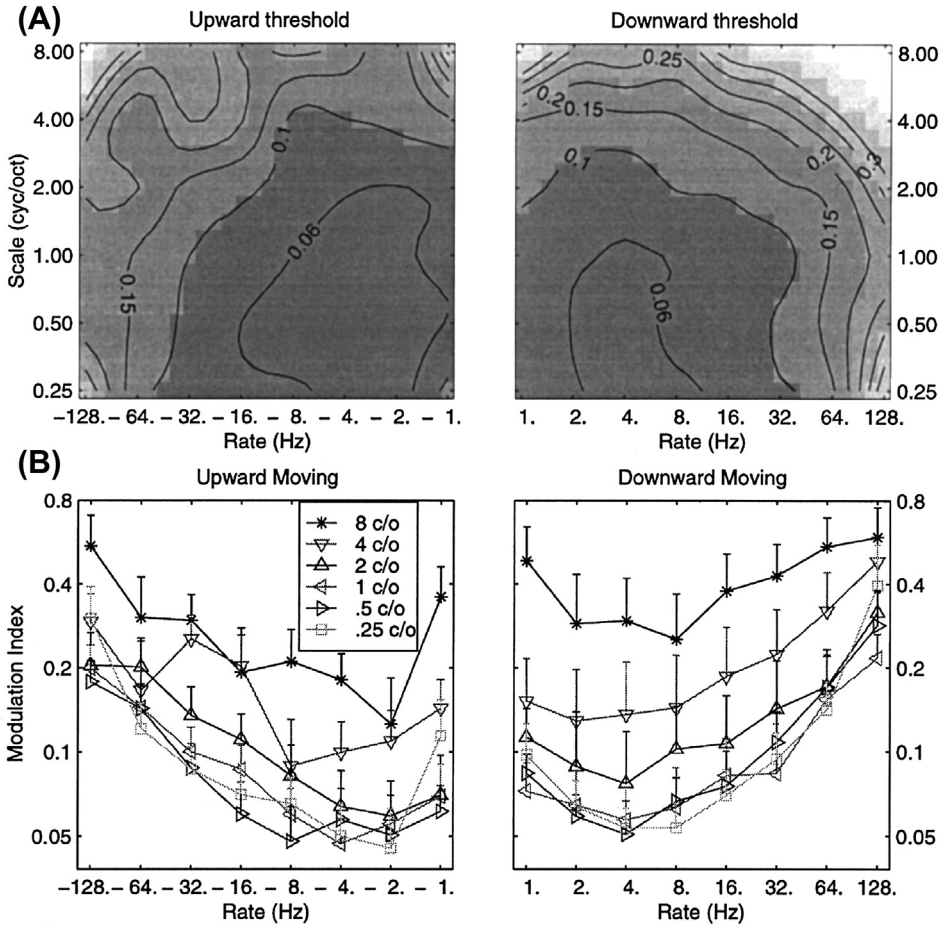

Fig. 7. The spectro-temporal modulation transfer function (STMTF) of Chi et al. (1999). In (A), the abscissa gives the temporal modulations (AM) in Hz, and the ordinate the spectral modulations (cyc/oct). Upward vs. downward ripple sounds (this concerns the orientation of the ripples in the spectrogram) are shown on the left vs. right of the figure, respectively. In (B), the temporal results are shown for various spectral frequencies, and overall exhibit a strong bandpass characteristic, centered at ~2–5 Hz.

3.5. *The spectro-temporal modulation transfer function (STMTF)*

A generalization of the TMTF is the spectro-temporal modulation transfer function (STMTF). Recall that the TMTF is defined by computing the envelope spectrum of input and output, and computing the transfer characteristic. Computation of the envelope spectrum is essentially a 1-D Fourier transformation of the envelope (followed by smoothing). In a psychoacoustic context (Section 3.4), "the TMTF" is obtained by using AM detection thresholds as the output variable. Computation of the STMTF is essentially a 2-D Fourier transformation of the spectrogram (followed by smoothing), and using psychoacoustic detection thresholds as an output measure. The sounds used in this case are 'ripple sounds', which vary on a continuum from purely temporal modulation (AM sounds) to purely spectral modulation (essentially harmonic stacks as in periodicity pitch). We do not cover all ripple-sound studies here, just the two most important historical works where the psychoacoustic methods were introduced (van Zanten and Senten, 1983; Chi et al., 1999), and one recent study of high interest for speech perception (Elliott and Theunissen, 2009).

1. The STMTF was introduced into psychoacoustics by van Zanten and Senten (1983), using two subjects (themselves), 2IFC, and a fixed spectral modulation frequency of 200 Hz (i.e., this is similar to a periodicity pitch sound with fundamental of 200 Hz). They found a peak sensitivity at a temporal modulation of ~1 Hz, declining monotonically above and below (i.e., a bandpass characteristic, although the reason for the peak at ~1 Hz is not clear). Since they only tested one spectral modulation frequency, this is not actually a study of the STMTF, and more recent studies are to be considered a major improvement.
2. Chi et al. (1999) were the first to study the full STMTF for ripple sounds, using 4 subjects, 2IFC, and a range of spectral modulation frequencies. The task was to choose the interval containing the modulated vs. unmodulated sound. Given the historical importance, their main results are shown in Fig. 7. Notice the clear bandpass characteristic in the vicinity of ~2–5 Hz for all spectral modulation frequencies, and for upward and downward oriented ripples. Importantly, they demonstrated that the spectral and temporal results are *separable*. That is, the matrix of numbers plotted in part (A) of their figure, can be decomposed by singular value decomposition, and the first component alone explains over 85% of the variance. This has several implications, one of which is that the temporal and spectral MTFs can be studied separately, and then simply multiplied together to form the 2-D STMTF to a first approximation. Thus, the many results cited above for purely temporal studies remain essentially valid in the spectro-temporal framework. It is therefore not surprising that the ~2–5 Hz bandpass characteristic was found here as it was in classic temporal studies of AM tones and noise. This result is also consistent with the fact that the same bandpass characteristic

has been found for all technical stimuli tested (AM and FM tones, AM white noise, etc.).

3. Elliott and Theunissen (2009) used manipulations of the spectro-temporal modulations of speech to study the effects of different regions on speech intelligibility. This is an important update to the studies of Drullman et al. (1994a,b), who manipulated the temporal envelope only. With notch filters (see their Figs. 5 and 6), the results show the strongest degradation of speech intelligibility in the range ~2–7 Hz. For low-pass filtering, the range was shifted somewhat higher (note that certain consonants require modulation frequencies in the range ~8–16 Hz for intelligibility, but these are isolated bursts/onsets of fast AM, not repetitive modulation). In any case, their results definitely do not support a purely low-pass view, since the temporal modulations below ~1 Hz made little contribution to intelligibility.

Overall, the results concerning spectro-temporal modulations confirm the classic bandpass results for AM sensitivity, and for relevance to speech. We note again that the bandpass characteristic is broadly peaked in the range ~2–5 Hz, and not a sharp peak declining rapidly to 0 on either side. Measurements must be made well below $f_m = 2$ Hz in order to clearly see the high-pass portion of the curve. It is surprising to us how consistent the frequency range of ~2–5 Hz is found. There has been some confusion recently as to which AM frequency range to cite for "theta" interest in speech (see commentary by Obleser et al., 2012). Although various psychoacoustic studies show (broad) peaks at 2–5 Hz, 2–3 Hz, 3–5 Hz, or 1–7 Hz, we believe that the best summary range (using integers) is decisively "~2–5 Hz". For brain wave research, this overlaps the "delta" (~1–4 Hz) and "theta" (~4–7 Hz) ranges, so in those contexts one might refer to "delta/theta" tuning. All of these are part of the "fluctuation" range (~1–10 Hz) as currently defined (Fig. 1). The full fluctuation range encompasses the delta, theta, and alpha (~7–14 Hz) bands of the EEG.

We now take the ~2–5 Hz peak in modulation sensitivity as an empirical finding which requires explanation, and turn next to neurophysiological studies.

4. Human neurophysiology

Human auditory EEG, MEG, PET, and fMRI studies of modulated sounds focus overwhelmingly on the periodicity pitch and roughness ranges. Scalp EEG and MEG studies in the 1980s and 1990s, later followed by fMRI studies, focused on the responses to 40-Hz repetitive or AM stimuli. This was driven initially by clinical and basic research interest (Galambos et al., 1981; Sheer, 1989), and then by the post-Singer (1992) interest in synchrony and 40-Hz. Overwhelmingly, these studies only included AM rates down to 5–20 Hz. The focus in human neuroscience on the faster AM rates is part of the historical reason that the viewpoint became dominant in the 1980s and 1990s that the TMTF is simply low-pass in nature (as would be observed trivially if the lowest AM rate tested is 5–20 Hz). Nonetheless, we identify a handful of fMRI studies providing evidence for the ~2–5 Hz bandpass characteristic.

The fMRI evidence is discussed first, because it is more straightforward in its interpretation. But even here there is one brief caveat of interpretation: If a given voxel is found to exhibit peak activation for some f_m, say 5 Hz, this does not mean that all neurons within the voxel have best modulation frequencies (BMFs) peaked at 5 Hz. It means that some weighted average over the neurons within the voxel yields a peak at 5 Hz. Also note that the BMFs do not necessarily reflect local cortical processing, but may be inherited from lower CNS processing. We will find that BMFs in the range ~2–5 Hz are unlikely to be inherited from lower brainstem centers, but this does not exclude the thalamus. Despite these caveats, we know that the psychophysical outcome depends ultimately on the population-level cortical activity, and fMRI yields roughly a measure of population-level firing rate, so evidence from this method should be useful with respect to psychoacoustics.

4.1. Human fMRI

Given the sluggish response of blood flow to cortical activation, fMRI is not used to measure cycle-by-cycle responses to modulated sounds (Harms and Melcher (2002) estimate ~0.1 Hz as the upper limit for fMRI). Typically, a given modulation rate (f_m) is presented for multiple seconds and the total blood-flow response is measured. Conceptually, this is a simple means of obtaining a tuning to modulation rate for a given brain region: present various f_ms and measure the total activation as a function of f_m. A number of studies of this type have been published, and a smaller number touch upon the ~2–5 Hz region of interest here.

As introduction to the fMRI evidence, we examine the now-classic study of Giraud et al. (2000), which is also noteworthy for discussing syllable-rate f_m (~2–5 Hz). They used SAM white noise at rates of 4, 8, 16, …, 256 Hz, and several basic facts about temporal processing are established here. A common set of brain regions were found to respond to modulated > unmodulated noise: these included the known subcortical auditory stations, Heschl's gyrus (HG), superior temporal gyrus (STG) and sulcus (STS), and supramarginal gyrus (SMG). It was noted by Scott et al. (2006) that essentially the same regions which respond greater to intelligible speech vs. speech-envelope-modulated noise, also respond greater to modulated vs. unmodulated sounds. That is, the STG/STS regions (homologous to monkey parabelt regions which respond to species-specific vocalizations) are activated by speech > AM noise > noise > silence. This shows the general relevance of AM sounds for speech perception regions.

Giraud et al. (2000) also confirm the general principle from animal neurophysiology (Section 4.2) that the best modulation frequencies (BMFs, i.e., the f_ms which elicit the strongest response) decrease with progress along the auditory pathway from cochlear nucleus (CN) to cortex. The CN responds to ~periodicity-pitch range AM, the inferior colliculus (IC) to ~roughness range, and the cortex to ~fluctuation range AM. The majority of AM-sensitive cortex showed the greatest activation to $f_m = 4$ Hz, their lowest frequency tested. They also found transient responses to $f_m > 16$ Hz, which were significant in a more restricted region (in or near the HG).

With these general points in mind, we look at the handful of fMRI studies which used low-frequency AM rates and provide evidence that the ~2–5 Hz psychophysical tuning is reflected in cortical activation:

1. The first auditory fMRI studies were published by Binder et al. (1994a,b). Binder et al. (1994b) used a presentation rate of 3 Hz in order to elicit strong activation, as did other early studies. Binder et al. (1994a) studied the effect of repetition rates, using syllables presented at 0.17–2.5 Hz. According to the band-pass perspective, this should reveal the high-pass portion of the curve (i.e., below the peak at ~2–5 Hz). Their results are consistent with this, showing monotonically increasing activation from 0.17 to 2.5 Hz in the superior temporal auditory regions. Frith and Friston (1996, using PET) studied tone repetition rates from 0 to 1.5 Hz, and also found a high-pass characteristic in superior temporal cortices. Rinne et al. (2005, fMRI) studied repetition rates of harmonic tones (periodicity-pitch stimuli) in superior temporal cortex. They found $0.5 < 1 < 1.5 < 2.5 < 4$ Hz or $0.5 < 1 < 1.5 < 2.5 = 4$ Hz,

depending on the state of (intermodal) attention. This is consistent with a high-pass characteristic below the peak region of ~2–5 Hz.

2. A number of fMRI studies (e.g., Giraud et al., 2000; Seifritz et al., 2003) looked at AM sounds or repetition rates in the roughness range (sometimes for interest in 40 Hz), with the lowest rate tested in the range 4–20 Hz. These studies found a low-pass characteristic, as expected for the range above the peak at ~2–5 Hz.
3. Tanaka et al. (2000) presented a 1-kHz tone (30-ms duration) at rates of 0.5, 2, 5, 10, and 20 Hz (each rate in separate 30-s blocks): "On the whole, the number of activated pixels increased up to a rate of 5 Hz and then decreased." This bandpass characteristic was statistically significant and is evident in their Fig. 4 (number of pixels activated) and Fig. 5 (percent signal change). This increase in activation strength *and* extent at 5 Hz is consistent with the results of Giraud et al. (2000) at 4 Hz (Fig. 8).
4. Harms and Melcher (2002) studied the IC, MGB, HG, and STG with white noise bursts at 1, 2, 10, 20 and 35/s (each rate in separate 30-s blocks). The peak activations were found at IC: 35/s; MGB: 20/s; HG: 10/s; and STG: 2/s. The decreasing rate preference with progress along the auditory pathway is consistent with Giraud et al. (2000) and with animal evidence (Joris et al., 2004; Malone and Schreiner, 2010). Keep in mind for this and other studies that the exact peak is only with respect to the coarse spacing between rates tested (so the peak at 2 Hz here is relative to 1 and 10 Hz). A nice methodological feature in this study was controls for intensity and for total intensity in a block, and such intensity effects were not found to drive the response (the cortex is sensitive to AM, but insensitive to overall amplitude).
5. Langers et al. (2003) studied ripple sounds in a 2IFC task (similar to Chi et al., 1999, Section 3.5) including temporal modulation frequencies of 2, 8, and 32 Hz. There was greater activation extent and level, particularly postero-lateral to HG, with $2 > 8 > 32$ Hz. Given the course spacing, this is consistent with either a low-pass or band-pass characteristic, but this study is noteworthy here because they confirmed the *separability* result of Chi et al. (1999). That is, not only is psychoacoustic sensitivity separable into spectral and temporal modulation transfer functions, but also apparently the cortical activation patterns. On the other hand, Schönwiesner and Zatorre (2009) report a lower degree of separability for ripple sounds using fMRI. Their temporal MTFs exhibited peaks between 2.8 and 3.7 Hz depending on ROI.

Thus, the overall psychoacoustic findings are basically confirmed here, such as the decreased sensitivity in the roughness range (Fig. 3) and increased sensitivity in the fluctuation range. We identify only 2 studies (Tanaka et al., 2000; Harms and Melcher, 2002) which included an adequate range of repetition rates to disclose the bandpass characteristic near ~2–5 Hz, and both studies essentially confirm this characteristic. Studies of slower repetition rates (below 2.5 Hz) for syllables (Binder et al., 1994a), simple tones (Frith and Friston, 1996), and complex tones (Rinne et al., 2005) are compatible with the high-pass portion of the curve, and a number of studies using faster AM or repetition rates are compatible with the low-pass portion of the curve (e.g., Giraud et al., 2000; Langers et al., 2003; Seifritz et al., 2003).

The processing of modulations in the fluctuation range and the sensitivity centered at ~2–5 Hz are more strongly associated with non-primary auditory cortex – regions surrounding HG, and particularly STG and planum temporale (PT) regions lying posterior and lateral to HG. Thus, modulations near ~2–5 Hz elicit not only greater *levels* of activation, but also a greater *extent* of activation given the larger size of non-primary vs. primary cortex. Boemio et al. (2005) also emphasized the role of belt/parabelt regions in "temporal structure" processing (see Fig. 11 for introduction to 'core' vs. 'belt' and 'parabelt'). Note that HG participates in the ~2–5 Hz finding, but it also expresses tuning to higher modulation rates in the 'flutter' range (~16 Hz). Hall (2005), who studied 5 Hz AM and FM stimuli (Hall et al., 2002; Hart et al., 2003), reached similar conclusions: "The results from the fMRI studies in humans converge on the importance of non-primary auditory cortex, including the lateral portion of HG (field ALA), but particularly subdivisions of PT (fields LA and STA), in the analysis of these slow-rate temporal patterns in sound." These regions involved in fluctuation-range AM/FM overlap heavily the regions involved in speech processing (Giraud et al., 2000; Scott et al., 2006), confirming again the relevance of modulated sounds to speech.

4.2. Scalp EEG and MEG

Compared to the relatively straightforward interpretation of fMRI studies, the interpretation of scalp EEG and MEG studies is extraordinarily difficult. First, one must clearly distinguish between spontaneous brain rhythms and stimulus-driven rhythms. Second,

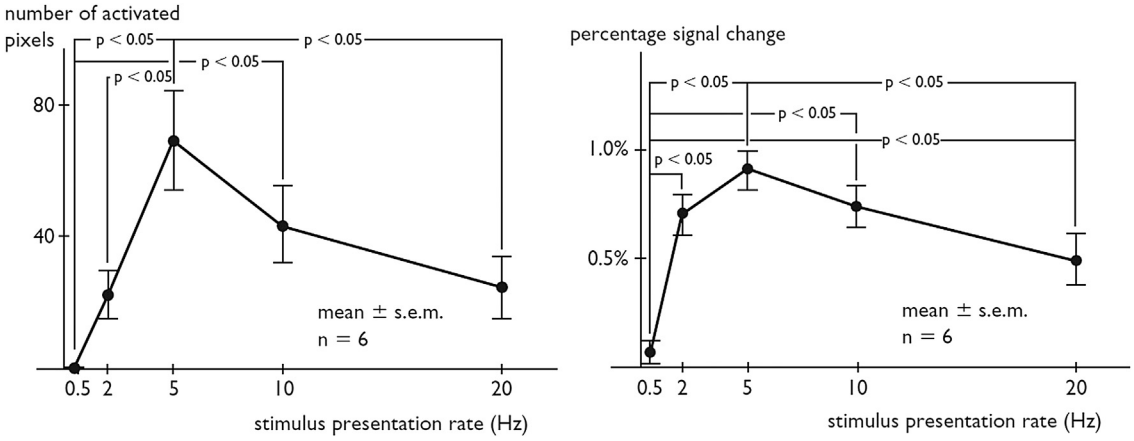

Fig. 8. fMRI results of Tanaka et al. (2000), using a sinusoidally-AM tone. Note that both the activation *extent* (left) and *strength* (right) increase near ~5 Hz. This figure concerns primarily HG, and can be compared favorably with monkey single-unit results below (Malone et al., 2007) (Section 5).

the biophysical and physiological origins of the signal are poorly understood, and, even if understood, the ability to localize the source of the signal is blurred by the skull. Any given EEG electrode effectively averages over primary and secondary auditory cortices, and non-auditory cortices. Third, there is a distinction between 'evoked' and 'induced' activity, and for the steady-state response (SSR) or AM situation we must decide on how to treat the event-related potentials (ERPs). The tuning to 40-Hz for the SSR, for example, is driven by the simple fact that the auditory middle-latency auditory evoked potentials (MAEPs) have their major peaks (Po–Na–Pa–Nb) separated by ~12.5 ms. That is, with 40-Hz AM or repetitive stimulation, the successive peaks overlap so as to reinforce each other. Now, this does not mean that the 40-Hz result says nothing about the time-scales of cortical processing, because it is probably not pure coincidence that the cortical ERPs exhibit this particular time separation. But it does make the interpretation of auditory SSRs more difficult, and particularly for slower AM rates where the long-latency auditory evoked potentials (LAEPs) will begin to overlap. This is probably one of the reasons why slower AM rates are rarely reported in the scalp EEG and MEG literature.

Another issue of interpretation, relevant also to fMRI, concerns the distinction between the *rate* TMTF (rTMTF) and the *synchrony* or *vector-strength* TMTF (vTMTF). In the rTMTF, the amount of modulation in the stimulus at some f_m is correlated with the total increase in firing rate in the neural response. In the vTMTF, the amount of stimulus f_m is correlated with the amount of modulation in the neural response at that same f_m. Clearly, the fMRI response is driven by the rTMTF, since the BOLD signal follows the total firing rate over the recent past of several seconds. In contrast, the raw EEG signal should follow the vTMTF, since oscillations of the EEG essentially follow the oscillations of pyramidal cell dipole strength and polarity (with an additional LTI system representing the extracellular transfer to the electrode). This is beyond the present scope, but is in line with the long held view that the EEG is driven by synchrony of synaptic potentials. The word "synchrony" here should not invoke any great mystery – this statement means nothing more than: the modulation at frequency f_m of the number of apical vs. basal synaptic potentials results in the appearance (with some LTI phase-shift and gain) of the frequency f_m in the raw EEG signal. This is standard dipole theory of EEG, but is beyond the present scope.

Thus, the interpretation of EEG results is not straightforward, and would require a separate full-length review alone. We only note that the work of Picton, a leading expert on auditory SSRs, shows results consistent with the predicted band-pass characteristic (Picton et al., 1987) and with clear relevance to the speech envelope (Aiken and Picton, 2008). However, there is inter-subject (and inter-electrode) variability and one can find confirmation of probably any perspective in the total EEG/MEG/ECoG literature. There are a large number of studies, some of them excellent and clear-headed in their interpretations, focusing on AM rates above the fluctuation rate, usually on or around the famous 40-Hz (Galambos et al., 1981; Sheer, 1989; Picton et al., 2003), but these are not relevant to the present focus on the fluctuation range. Overall, we are not able to draw and strong conclusions from the scalp EEG/MEG literature for fluctuation range AM/FM.

In order to complete our task of usefully bringing together basic information about the fluctuation (~1–10 Hz) range, we include Fig. 9. Since there has been critique recently concerning the inconsistent use of different Greek-letter frequency bands (δ, θ, α, β, γ), this figure reviews the historical introduction and current conventions for these symbols, although the gray transitions indicate an acceptable sloppiness for the boundaries. Note that these are merely useful labels, although the ranges do correspond to some degree to distinct categories of spontaneous brain rhythms; critics

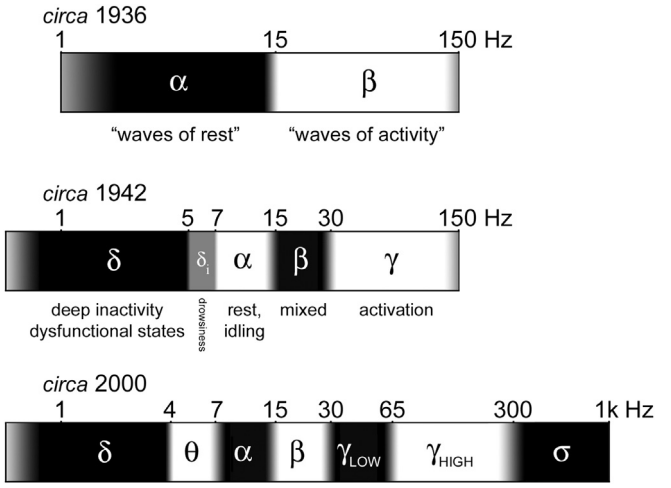

Fig. 9. History of the EEG spectrum. All boundaries are approximate (indicated by gray transitional areas), due to different usages by different authors, but best-fit integers for the boundaries are indicated. The frequency scale is logarithmic. Top (*circa* 1936): Two frequency bands, α and β, were distinguished by Berger (1930), with the division at ~15–20 Hz. The lower/upper boundaries of α/β were not specified (frequencies outside the range ~2–150 Hz were not studied then). Ectors (1936), who systematically mapped sensory and motor cortices in awake rabbits, aptly referred to these ranges as "waves of rest" and "waves of activity". Middle (*circa* 1942): δ (~0.5–5 Hz) was introduced by Walter (1936) for slow rhythms from dysfunctional tissue in the vicinity of tumors, and was quickly adopted for slow waves during deep sleep (Davis et al., 1937). γ (above ~30 Hz) was introduced by Jasper and Andrews (1938), although it had been discovered and systematically used for brain activation mapping by Ectors (1936). The "intermediate δ" band (~5–7 Hz) was introduced by Jung (1941) as a sign of drowsiness, and is included here as the precursor to today's θ band (~4–7 Hz), also well-known to correlate with drowsiness. Bottom (*circa* 2000): The contemporary EEG spectrum includes the θ band (Walter and Dovey, 1944) and division of the γ band into low (~30–60 Hz) and high (~65–300 Hz) regions (Crone et al., 1998). A set of phenomena above ~300 Hz are generated by summed multi-unit spiking, labeled here the σ band (after Curio, 2000). Note that the recent labels (high-γ, σ) are not universally accepted, and note that high/low divisions have been proposed for other bands also.

since Grass and Gibbs (1938) have emphasized that the EEG spectrum is to be thought of as a continuum. This is understood by all workers in the field and there is nothing wrong with using these labels as quick reference, so long as the usage is clear.

We note that the classic view that slow rhythms (δ, θ, α) represent cortical inactivity or 'idling', whereas fast rhythms (γ) represent cortical activation (Ectors, 1936; Pfurtscheller, 1999; Crone et al., 2001), has been abundantly confirmed by blood-flow and metabolic measures (Darrow and Graf, 1945; Logothetis et al., 2001; Mukamel et al., 2005). Thus, future work on the role of "θ" in speech EEG/MEG should be careful to distinguish spontaneous θ, which is generally a sign of drowsiness and idling, from stimulus-driven θ. For example, Scheeringa et al. (2009) recently demonstrated with simultaneous EEG/fMRI that certain θ increases, thought to be due to cognitive *activity*, were in fact just increases in cortical *idling* in non-engaged parts of the cortex. Thus, researchers using scalp EEG/MEG to study stimulus-related and top-down θ influences during speech must proceed with particular caution in methods and interpretation.

This is not to discourage work in this area, and we believe that important missing evidence will be provided by EEG/MEG (or perhaps LFP/ECoG) studies. For example, Kenmochi and Eggermont (1997) report a correlation between a cortical neuron's BMF and frequency of spontaneous oscillation in the LFP of cat auditory cortex. They also report correlations between click-following rate and the amplitudes of local idling rhythms. The *slower* regions of cortex (in terms of click-following rate) exhibit *larger* spontaneous rhythms. It has been known since the 1940s LFP/ECoG literature

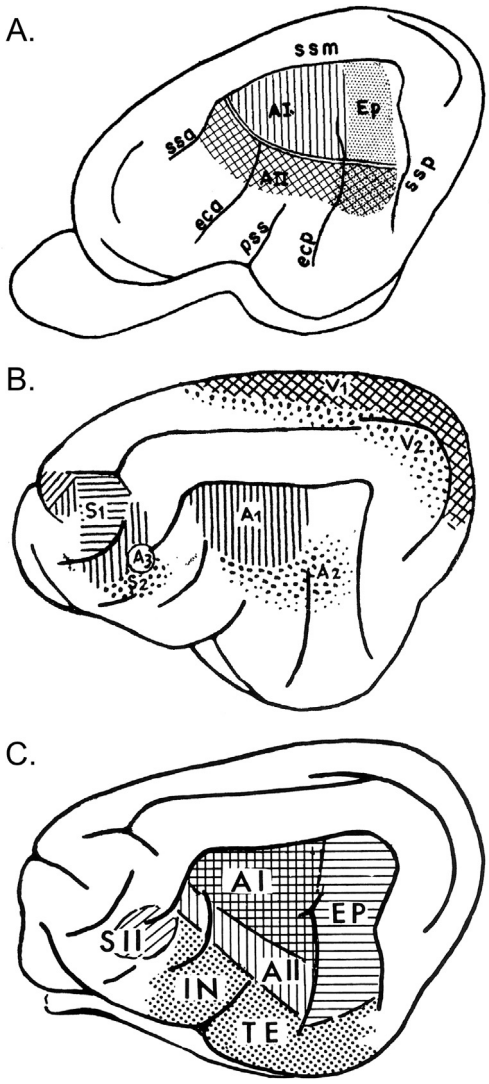

Fig. 10. Cat cortex: quick orientation and terminology. A. The basic sensory-responsive cortices from Bremer (1952). 'A3' is a small auditory-responsive zone within S2 (multimodal). B. Core (AI) and belt (AII, Ep) regions from Rose and Woolsey (1949), using anatomy and evoked-potential mapping ('Ep' = posterior ectosylvian area; 'ss' = suparasylvian sulcus). C. Summary of auditory-responsive regions from Ades (1959) including core (AI), belt (AII, EP), and parabelt/association ('IN' = insular region; 'TE' = temporal area) regions. Note that these regions, based on ECoG evoked-potentials, are expanded (spatially blurred) compared to current maps based on single- or multi-unit mapping.

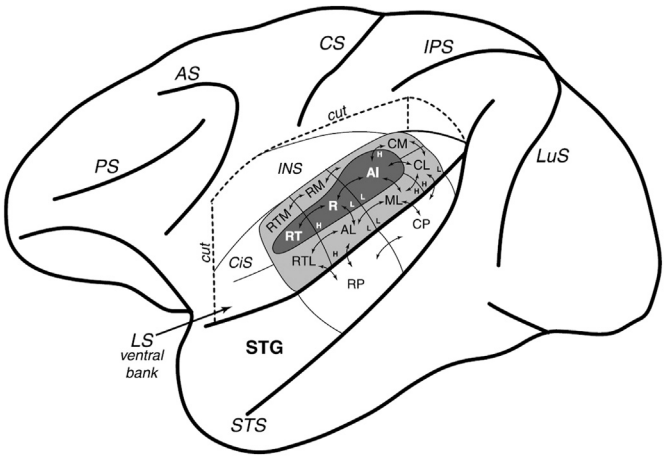

Fig. 11. Macaque cortex: core, belt, and parabelt regions from Hackett et al. (2001) (who also studied humans). 'Core' regions (AI, R, RT) are in dark gray within the lateral sulcus (LS, e.g. the 'Sylvian fissure'). 'Belt' regions (CM, RM, RTM, RTL, AL, ML, CL) are in light gray. 'Parabelt' regions (RP, CP) occupy the major exposed surface of the superior temporal gyrus (STG). Note that human anatomical organization is suggested to be similar (Hackett et al., 2001; Sweet et al., 2005; Fullerton and Pandya, 2007; Brugge et al., 2008; Baumann et al., 2013).

that non-primary regions exhibit larger spontaneous rhythms compared to primary regions, which is therefore in line with their slower BMFs during auditory stimulation (Section 5). Thus, the LFP/ECoG spontaneous rhythms may reflect the temporal dynamics observed during stimulation.

5. Animal neurophysiology

Based on the human fMRI evidence, two expectations for (unanesthetized) primate cortex are: 1. Primary regions exhibit some ~2–5 Hz AM/FM tuning, but also tuning up to ~20 Hz or more, with the peak at ~5–10 Hz. 2. Non-primary regions, particularly those lying lateral to HG, are generally slower and appear to more strongly express the ~2–5 Hz peak AM/FM tuning.

Several excellent reviews exist for general AM/FM results in animal neurophysiology (Kay, 1982; Langner, 1992; Joris et al., 2004; Wang et al., 2008; Malone and Schreiner, 2010), and our main purpose here is to understand results from humans, so we do not give a comprehensive review. However, we can still provide a useful focus on the fluctuation range (~1–10 Hz), and attempt to determine the neural correlates of the observed psychophysical tuning to AM/FM with broad peak at ~2–5 Hz. The fMRI results predict a bandpass tuning in the *population-level firing rate* with a broad peak at ~2–5 Hz in certain non-primary auditory cortices. Since the prediction from human fMRI focuses on (lateral) non-primary regions, understanding of the evidence requires a brief introduction to *core* vs. *belt* regions.

5.1. Basic orientation: core vs. belt

This basic distinction for auditory cortex emerged in the 1940s in anatomical and physiological studies of the cat (Fig. 10), and reviews by the key early workers can still be used for basic orientation (Rose and Woolsey, 1958; Ades, 1959; Woolsey, 1961). The early physiological workers used evoked potentials (ECoG), which resulted in blurred spatial resolution compared to more modern maps (e.g., Fig. 11) based on multi- or single-unit spike rates (Merzenich and Brugge, 1973; Imig et al., 1977; Aitkin et al., 1986). A *tour de force* history is given by Jones (2010). The 'core' vs. 'belt' distinction is also extended to *subcortical* structures (Andersen et al., 1980; Calford and Aitkin, 1983; Aitkin, 1986), which is essential to our hypotheses concerning the origins of the ~2–5 Hz tuning, so it will be illustrated below (Fig. 14).

Comparing cats and monkeys for homologous regions is not always straightforward, but at least AAF and CM appear established as near-homologs. There has been a huge transformation of the neocortex (expansion of association areas, greater number and depth of sulci) going from carnivores to primates. The peri-Sylvian auditory regions appear rotated by nearly 180°, hence an anterior field matching to a caudo-medial field. AAF/CM is the best studied auditory field outside of the core, and is in some ways more similar to core than to belt areas (Imaizumi et al., 2005). It tunes to higher modulation frequencies than even AI, and so AAF/CM is to be excluded from certain summary statements about 'belt' regions, such as their typically slower characteristics. As a final point of general orientation (Rauschecker and Tian, 2000), caudal belt areas

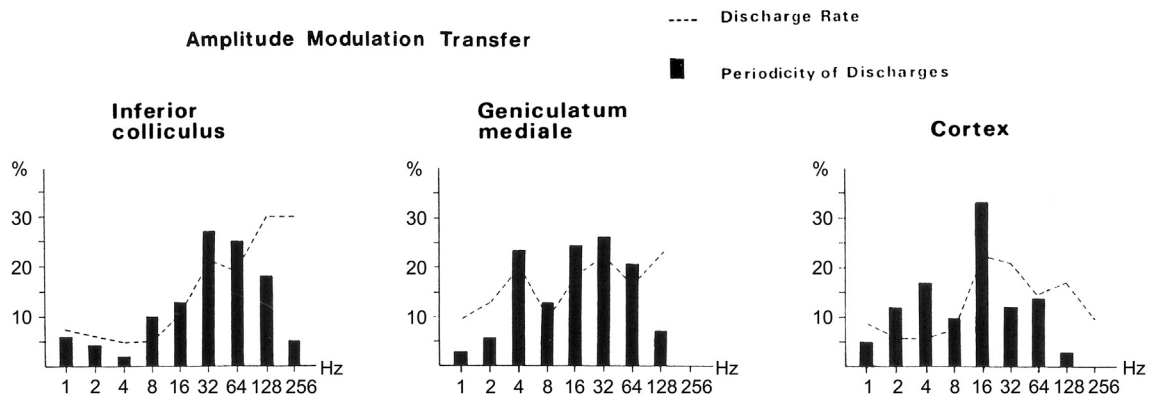

Fig. 12. Best-modulation frequencies for AM sounds in IC, MGB and auditory cortex of awake primates (Müller-Preuss et al., 1988) (the axes have been relabeled with increased font-size). Note the drastic increase in ~4 Hz tuning at the MGB and cortex compared to the IC. The IC and cortex data is from the core regions only, whereas the thalamic data mixes core and belt regions.

(CL, and sometimes CM) are implicated in an auditory 'where' pathway (toward parietal lobe regions for spatial processing), whereas lateral belt regions begin a 'what' pathway (toward temporal lobe regions for processing of natural sounds and species-specific vocalizations). The evidence for functional specialization does not appear unequivocal to us, and we mention this only as a point of general orientation: In these terms, the fMRI evidence predicts slower (~2–5 Hz) tuning in belt and parabelt regions of the 'what' (lateral) pathway.

For humans (Hackett et al., 2001; Sweet et al., 2005; Fullerton and Pandya, 2007; Brugge et al., 2008; Baumann et al., 2013): the core is localized to HG, the belt to surrounding regions of the supratemporal plane (STP) and lateral HG, and the parabelt to further surrounding regions, including most of the exposed surface of the superior temporal gyrus (STG). Lateral belt regions may just emerge from the Sylvian fissure onto the exposed STG. The fMRI activations for fluctuation-range AM/FM were found most strongly in regions lateral to HG, which are belt and parabelt regions. HG displayed some ~2–5 Hz tuning, in addition to faster tunings (up to ~20–32 Hz), so 'core' regions of animal cortex should exhibit a subset of cells with this characteristic.

5.2. Single-unit studies of core cortex (AI)

Single-unit studies of auditory cortex began in the 1950s (Erulkar et al., 1956), but the early studies were concerned overwhelmingly with methodological issues and basic response properties to clicks and tones (latency, intensity relations, tonotopy, etc.). Katsuki et al. (1960) briefly mention "remarkable" responses to beating tones in unanesthetized monkeys, but no specifics are given. Some early animal ECoG studies (Goldstein et al., 1959) used repetitive click stimuli with focus on periodicity pitch, and later studies also focused on these rapid repetition rates, but these are not directly relevant to the present focus. Single-unit work of the 1970s ± a decade focused heavily on subdivision of cortical fields, tonotopy, and other response properties of AI. Thus, despite a long history of work on AI, we find only a small number of studies for fluctuation range AM/FM:

1. Whitfield and Evans (Whitfield, 1957; Whitfield and Evans, 1965; Evans, 1968) studied AI of unanesthetized cats using FM stimuli, including sinusoidal FM in the fluctuation range. Whitfield (1957) found that the ECoG (surface potential) from AI could be driven at the same rate as the FM for rates between 2 and 18 Hz. Whitfield and Evans (1965) found that most single-units responded to FM tones more consistently than to static tones. Testing a range of FM rates, they found: "Rates as low as 1 cycle/sec. were still effective in a few units in evoking periodic firing consistently related to some point on the modulation waveform. For most units, however, rates below 2–3 cycles/sec. and above 15 cycles/sec. tended to be less effective in evoking the consistent responses…" Evans (1968) discusses this result in terms of the emphasis in cortex on dynamic vs. static stimuli.
2. Fastl et al. (1986) searched for the neural correlates of fluctuation strength in AI of unanesthetized squirrel monkeys. SAM tones from 0.5 to 32 Hz AM were tested at various modulation depths, and the correspondence with human psychophysics was seen as "promising". Specifically, AI neurons generally exhibited bandpass characteristics with best modulation frequencies (BMFs) below 32 Hz, often in the upper fluctuation range (~5–10 Hz).
3. Müller-Preuss et al. (1988) studied IC, MGB, and auditory cortex in unanesthetized squirrel monkeys using SAM noise and tones. AM rates from 1 to 256 Hz were tested for ~450 units total. They confirm Fastl et al. (1986) in that: "the most impressive result is that most of the units are sensitive within a particular band of AM-frequencies. There are only a few units which display a low pass characteristic or have complex response patterns (i.e. multiple peaked)." Their full data for IC and MGB will be discussed in Section 5.4, but here we note the appearance of a peak near ~4 Hz (for vTMTFs) in the thalamocortical data compared to IC (Fig. 12). However, the full report of the cortical data (Bieser and Müller-Preuss, 1996), with more extensive measurements, shows a broad peak at ~8 Hz for core regions, not two peaks at 4 and 16 Hz. The majority of core BMFs were in the range 1–32 Hz, so the results are overall consistent with core results in other primates (although squirrel monkeys appear to exhibit overall faster AM tuning preferences than Old World primates, Brian Malone, personal communication).
4. Eggermont (1993, 1994) studied AM noise and AM/FM tones in lightly anesthetized cat (light ketamine, and he provides some evidence that the anesthesia does not drive the results, so they are included). In AI of the adult cat, synchrony-BMFs for AM noise peaked in the range 8–12 Hz, whereas for AM/FM tones they peaked mostly in the range 4–7 Hz (full range up to 32 Hz or more). Eggermont heavily studied click trains, which give results most similar to AM noise (both are broadband stimuli), but we do not cover this here. See also Eggermont (2002) for comparison of click trains to other AM/FM stimuli, where he points out that AI neurons prefer stimulus categories with rapid onsets (clicks, gamma tones, etc.).

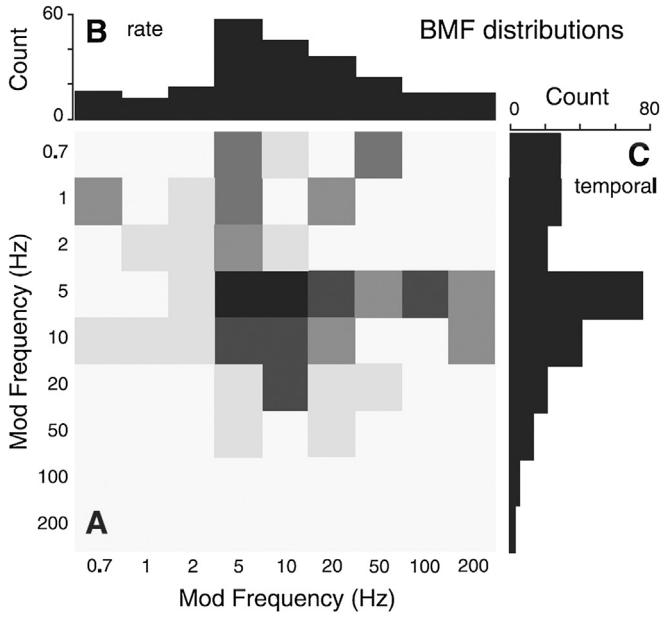

Fig. 13. Single-unit responses to SAM tones in AI of unanesthetized monkey (Malone et al., 2007). For each single-unit, a best modulation frequency (BMF) was calculated either based on overall firing rate increase (B) or on a temporal measure (C). Both measures indicate a peak at ~5–10 Hz, with the majority of BMFs found in the range ~4–32 Hz. The joint distribution (A, 361 AI neurons total) shows that the two measures are roughly, but not perfectly, correlated.

5. Liang et al. (2002) studied AI in awake marmosets, using SAM and SFM tones. Modulation frequencies as low as 1–4 Hz were used and most single-units exhibited a band-pass preference. The majority of rate and synchrony BMFs were in the range 4–32 Hz, although rate BMFs in particular can be found as high as 128–256 Hz. They emphasize the similarity in their results for AM and FM stimuli, although we note slightly lower BMFs for SFM (see similar comment in Section 3). Bendor and Wang (2008) showed that, within the core cortical regions of awake marmoset monkeys, R and RT have longer latencies and slower AM tunings than AI. They proposed a caudal-to-rostral gradient of increasing temporal integration with implication of hierarchical progression.
6. Malone et al. (2007) studied SAM tones in core regions (AI, R) of unanesthetized macaques, and found the majority of BMFs in the range ~4–32 Hz (peak at ~5–10 Hz for their f_ms tested). They emphasize the lack of correlation between rate and temporal BMFs, but the peaks/ranges are similar for rate and temporal BMFs (Fig. 13), so the population-level result is essentially similar.
7. Yin et al. (2011) studied AI in awake macaques using SAM noise (the lowest rates tested were 5 and 10 Hz). The great majority of rate and synchrony BMFs were in the range 5–30 Hz, with the peak at 5–10 Hz.

Finally, there are a number of studies in AI of anesthetized animals (e.g., Depireux et al., 2001) which appear to generally confirm these tuning ranges for unanesthetized AI, but we do not cover these here. We only note that anesthesia can slow responses, so tunings may appear slightly downshifted in these preparations. We also note clear species differences for rats (much smaller cortex) and bats (specialization for echolocation).

Overall, the cat and monkey results in core auditory cortex are confirmatory of human fMRI findings for AM/FM tuning in HG: most BMFs are found between ~2–32 Hz, with the peak (most likely BMFs) at ~4–12 Hz. We do not find in the animal literature the plot which is really needed for comparison to human measures: the AM/FM tuning of the population-level firing rate. In the meantime, we must be content with a rough approximation to the desired confirmation, which is found in the distributions of single-unit BMFs.

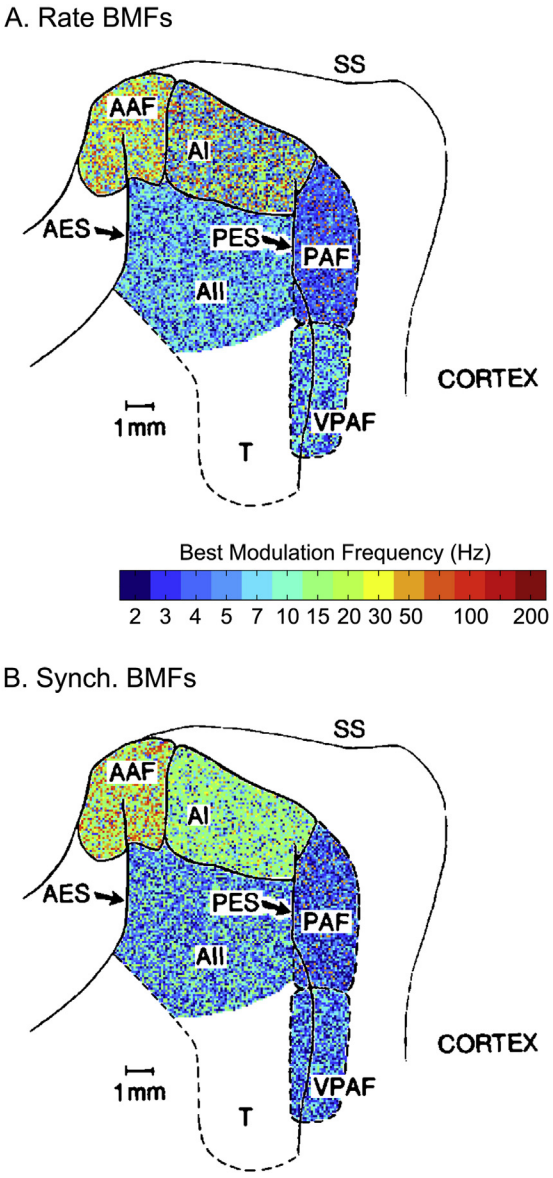

Fig. 14. Best modulation frequencies (BMFs) for AM tones as a function of anatomical region in the cat (Schreiner and Urbas, 1988). The underlying map is adapted from Andersen et al. (1980), whose terminology is employed by Schreiner and Urbas. Rate (A) and synchrony (B) BMFs were obtained for 172 single-units using 14 AM rates from 2.2 to 200 Hz (see colorbar, note that "2" really means "≤2.2" since no lower AM rates were tested). Each of the five cortical regions was colored according to the proportion of BMFs observed at each rate (positions within a given field are assigned randomly). This quickly summarizes the results for the main conclusions: AAF exhibits the fastest BMFs (up to 100 Hz, but still typically near ~20 Hz) followed by AI. The 'belt' regions AII, PAF, and VPAF exhibit the slowest BMFs, overwhelmingly in the fluctuation range (~1–10 Hz). PAF (receiving heavy input from MGBd) is the slowest, with a clear preference for ~2–5 Hz AM.

5.3. Single-unit studies of belt cortex

As mentioned above, single-unit work has focused overwhelmingly on core regions, or on tonotopy and delineation of

cortical fields when outside of the core. The review of Goldstein and Knight (1980), for example, noted how remarkably little work had been done on auditory belt regions. For fluctuation-range AM/FM, we find only 1 note of anecdotal evidence concerning belt regions prior to 1988, and only 4 studies since. We make careful survey of this evidence since it is central to our interpretation of human findings.

1. Galambos (1960) surveys his results in unanesthetized cats and reports units which respond well to FM "warble" at 2–3 Hz, but not to steady tones. He also reports broad-band units responding well to 3–8 Hz repetition rates, but poorly to constant stimuli. These units occur more often in belt area Ep rather than AI, but details are lacking.
2. Schreiner and Urbas (1988) studied single-units in lightly anesthetized cats from a range of primary and non-primary cortical fields. The stimuli were AM tones, with the carrier frequency set to the unit's CF, and AM rates from 2.2 to 200 Hz were tested. BMFs were computed both for total firing rate increase and for amount of phase-locking to the AM envelope ('synchronization'). The results are shown in Fig. 14, and clearly indicate higher BMFs in core (AI) than in belt (PAF, VPAF, AII) regions. AAF, like its monkey counter-part CM, displays even higher BMFs than AI (and will be excluded from summarizing statements about core vs. belt below). Although the results are confirmatory of expectation from primate and human studies, the use of anesthesia is of concern here despite claims (also from Eggermont) that light anesthesia should not greatly influence these types of results.
3. Bieser and Müller-Preuss (1996) studied awake squirrel monkeys (a small New World primate) in 8 different core and belt regions including insula, and concluded that: "The eight cortical areas investigated displayed clear differences in their ability to encode amplitude envelopes." The plotted results exhibit considerable variability, but can probably be said to confirm the overall expected patterns (noting again that New World monkeys exhibit tuning to somewhat higher AM rates). Specifically, the great majority of BMFs were in the range 1–32 Hz, confirming other cortical studies, and certain belt regions were clearly slower in AM tuning than core regions (namely, their 'AL', 'Pa', 'RPi', and insula). Their 'Pi', in a caudo-medial position, was similar to core regions (probably confirming results from AAF/CM). Lateral belt region T1 was partly slower, but also exhibited a strong peak at 16 Hz, which we speculate could be due to specialization for squirrel monkey vocalizations (the 'twitter' vocalization is centered near ~12 Hz, and they tested only octave AM spacings 1, 2, …, 8, 16, …Hz). A second explanation is that their T1 extends along the entire lateral border of core regions, so according to the caudal-to-rostral gradient of Bendor and Wang (2008), this should mix neurons with faster (caudal) and slower (rostral) AM preferences.
4. Eggermont (1998) studied AI, AAF, and AII of lightly anesthetized cats, and found that "similarities outweigh differences". However, the search stimuli were gamma tones with rapid onset, and the 'AM' stimuli were mostly click trains and AM noise with an exponentiated-sinusoidal envelope (this concentrates the energy near the peak of the modulating waveform, and thus lies between SAM noise and click trains). SAM tones were also used, but reported to be ineffective stimuli for AII. Recall that AAF is expected to be similar to AI, or faster in AM tuning, and that AII showed the least differences from AI amongst the ventral/posterior belt fields in cat. Thus, this study cannot be taken as a strong disconfirmation of Schreiner and Urbas (1988), although it does alert us to the fact that AII differences may not be easy to observe with other stimuli and recording conditions.
5. Scott et al. (2011) reported on SAM and SFM tones in 2 unanesthetized macaques, and reported overall similarity of core vs. belt responses. However, the 'belt' data that was similar to AI was largely from CM (their medial belt field 'M' overlaps CM). Their 'L' cells appear to lie in 'ML', or on the border between AI and 'ML', in terms of the map shown in Fig. 12. Examination of their data shows that 'L' was the slowest of the fields studied (longer latencies and lower synchrony BMFs for SAM tones). While all fields showed a peak at 5 Hz for percentage of cells exhibiting envelope synchrony, the nearest modulation frequencies tested were 2 Hz and 10 Hz. Even at this coarse spacing, the AM/FM tunings of 'L' are clearly distributed toward lower values compared to core and CM fields. Thus, their results may actually be taken as consistent with those in Schreiner and Urbas (1988) for lightly anesthetized cats (Fig. 14). They are also consistent with the caudal-to-rostral gradient of Bendor and Wang (2008), but since rostral fields were not tested, the slower AM tunings were not as frequently detected.

The evidence from belt regions is obviously too sparse for definite conclusions. The observation of Galambos (1960) is essentially anecdotal, and the report of Eggermont (1998) appears (weakly) contradictory to a difference between AI and AII in cat. Field AAF (monkey CM) consistently shows modulation tuning similar to or somewhat higher than AI. The most systematic data (Schreiner and Urbas, 1988; Bieser and Müller-Preuss, 1996), which clearly indicate differences in AM tuning for belt areas other than AAF/CM, are from lightly anesthetized cats and squirrel monkeys (small New World primates with possibly faster AM tunings), respectively. Only the recent data of Scott et al. (2011), from belt region 'L' in unanesthetized macaques (Old World primates), gives clear preliminary support for the prediction from human fMRI: lateral belt areas exhibit slower AM tuning (mostly in the fluctuation range ~1–10 Hz), compared to faster AM tuning (up to 32 Hz) in core regions and AAF/CM. Even for macaques, recent behavioral results are quite different for AM noise detection compared to humans (O'Connor et al., 2011), so perhaps the only relevant cross-species observation is that belt areas (other than AAF/CM) are generally slower (i.e., lower BMFs overall) than core areas within any species.

A second theme which appears consistently supported in the evidence to date is the rostral-to-caudal gradient of Bendor and Wang (2008), whereby AM tunings for slower modulation rates are found more rostrally both within the core (R < AI) and the belt (RTL/AL < ML < CL). The most evidence is available for CM, which responds with shorter latencies than AI and is tuned to higher AM rates, consistent with being the most caudal field. ML ('L' of Scott et al. 2011) appears to be tuned to somewhat lower AM rates than AI. RTL/AL exhibited the lowest synchrony BMFs ($\leq$ 8 Hz) of all the fields tested by Bieser and Müller-Preuss (1996) (their 'AL' and 'RPi' are 'RTL' and 'RTM' of Fig. 11). Additional evidence for the caudal-to-rostral gradient is found in a study of linear FM sweeps in rhesus monkeys (Tian and Rauschecker, 2004), who found: "Neurons in AL generally responded better to slower FM sweeps (in the range of tens of Hz/ms), whereas neurons in CL responded best to very fast FM sweeps (in the range of hundreds of Hz/ms). ML neurons included all FM rates, …" Both Tian and Rauschecker (2004) and Bendor and Wang (2008) interpret the slower responses of more rostral fields as consistent with a higher position in hierarchical cortical processing. The anatomical work of Kaas and colleagues in primates also emphasizes a rostral/caudal distinction for core, belt, and parabelt regions, with hierarchical implications (Kaas et al., 1999; Kaas and Hackett, 2000). Thus, another summary

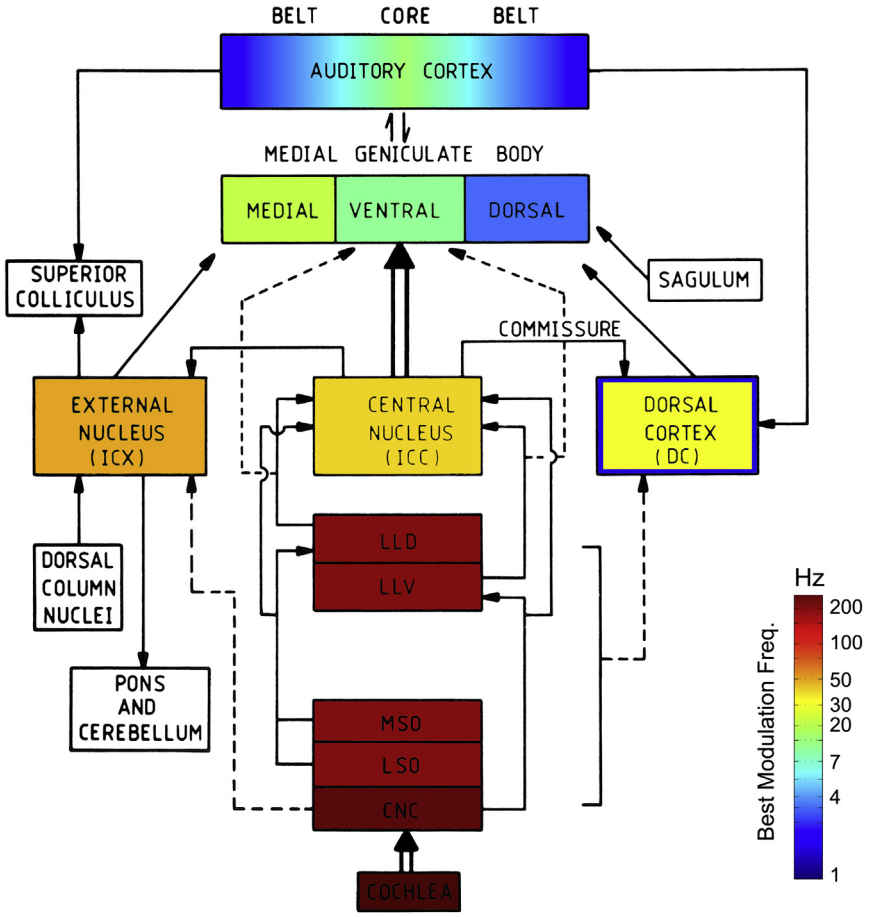

Fig. 15. Illustration of basic auditory CNS connections and typical BMFs (based on studies in various mammals, and likely applicable overall to humans). The basic diagram is adapted from Aitkin (1986), and the data for the BMFs is based primarily on Joris et al. (2004) for 'core' regions. The colors have been selected according to the overall central tendency and are for *illustrative purposes* only. For full quantitative results, consult Joris et al. (2004) and the references therein. The main purpose for the figure is to illustrate our own hypothesis (prediction) about BMFs in the 'belt' regions. There is little existing data for IC 'belt' regions (ICX, ICD), or for belt regions of thalamus (MGBm, MGBd). But, as argued in the text, ICD may give the slowest AM tunings of any IC subdivision, including a significant fraction of cells with low-pass or 1–2 Hz tuning (indicated by the blue rim). We also depict for thalamic BMFs that overall MGBm > MGBv > MGBd. The diagram shows our hypothesis that fluctuation range (~1–10 Hz) tuning arises primarily in 'belt' regions of cortex, although ICD and MGBd may be involved. In any case, the observed BMFs in the majority of the IC, and in lower parts of the auditory pathway, are entirely too fast (roughness or periodicity pitch ranges). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

statement for fluctuation range tuning is that it is associated with the higher levels of the auditory 'what' processing stream.

Overall (human and animal), the observed psychophysical AM/FM tuning (broad peak at ~2–5 Hz) appears most likely to be associated with lateral belt and parabelt areas, with a stronger weighting toward rostral regions, although core and other belt regions are not to be excluded as some neurons in these regions also show BMFs in the lowest AM ranges. On the other hand, these regions are heavily interconnected with further temporal, parietal, and frontal regions, so perhaps the final psychophysical outcome is only to be associated with a top-down signal from these higher-order regions (Dik Hermes, personal communication) or with the coordinated dynamics of these higher-order cortical regions interacting with belt/parabelt regions (Christoph Schreiner, personal communication).

5.4. *On the origins of fluctuation tuning in non-primary auditory cortex*

Although the observed psychophysical outcome depends more or less directly on cortical activity, it is possible that the AM response characteristics in cortex reflect preprocessing in the auditory periphery and brainstem. We briefly examine the evidence in this section and conclude that fluctuation (~1–10 Hz) or syllabic (~2–5 Hz) range tuning arises at the cortical or thalamocortical level, and not likely at any lower levels.

A general principle which has emerged from animal neurophysiology (Joris et al., 2004; Malone and Schreiner, 2010), and was confirmed by human fMRI (Section 4), is that BMFs decrease with progress along the auditory pathways. The highest AM tuning rates are observed in cochlear nucleus (CN), and the lowest in cortex (other than low-pass units found throughout the CNS). Joris et al. (2004) have reviewed and summarized evidence for the 'core' (or 'lemniscal') auditory pathway. We do not review the primary evidence ourselves and will instead concern ourselves with the 'belt' (or 'non-lemniscal') contributions. The ranges of observed BMFs in lemniscal centers are illustrated in Fig. 15. The first major conclusion for our purposes is that fluctuation range (~1–10 Hz) tunings are rarely observed below the level of the thalamocortical system. In fact, the thalamic and AI tunings are also faster than expected by simple interpretation of the psychophysical findings, leaving only the larger non-primary regions as the most likely candidate for a straightforward model. The only exception to this rule is possible contributions of non-lemniscal parts of the IC and thalamus, so these are discussed next.

The 'core' vs. 'belt' distinction for cortex has been extended to thalamus and IC (but not generally lower) (Andersen et al., 1980;

Aitkin, 1986). For the IC, the 'core' region is the large central nucleus (ICC), whereas the 'belt' region is a surrounding set of cells, sometimes referred to collectively as the 'pericentral', 'paracentral', or 'peripheral' nuclei. However, a more current terminology (Morest and Oliver, 1984; Irvine, 1986; Oliver, 2005) distinguishes within the 'pericentral' division at least the dorsal cortex (ICD) and the external cortex (ICX). These divisions are also distinguished in terms of their forward connectivity to the thalamus (Aitkin, 1986; Wenstrup, 2005). In the thalamus, the 'core' auditory nucleus is the ventral medial geniculate body (MGBv), whereas the 'belt' nuclei are the dorsal (MGBd) and medial (MGBm) divisions. Although all divisions of the IC project to some extent to all divisions of the MGB, the dominant projections are ICC → MGBv, ICD → MGBd, and ICX → MGBm.

Only the study of Müller-Preuss et al. (1994) reports AM results as a function of central vs. 'peripheral' divisions of IC. They report overall similar ranges of BMFs in both divisions (both peaking around 32–64 Hz), with the most noticeable difference in the lowest BMFs (1 Hz and 2 Hz). For peripheral nuclei, these represent ~20–30% of cells, whereas in central nuclei they represent only ~2–7% of cells (depending on rate vs. synchrony BMF measures). We can interpret these slower cells in terms of the classic 'periodotopy' result of Schreiner and Langner (1988) in cat, whereby higher BMFs are located centrally within ICC, and lower BMFs toward the external shell. If this trend is continued outward, then the peripheral nuclei should exhibit still lower tuning. However, the results of Müller-Preuss et al. (1994) do not support a simple continuation of a central-to-external periodotopy into *all* of 'peripheral' IC. We can, however, interpret their findings in light of the recent high-resolution fMRI study of primate IC (Baumann et al., 2011). Here a gradient of periodicity tuning was found such that ventro-lateral regions exhibited the fastest tuning (128 Hz or more) and dorsal-medial regions the slowest tuning (2 Hz or low-pass). If this gradient were to extend into the peripheral divisions, then ICD should exhibit lower BMF tuning and ICX should exhibit higher BMF tuning. Inclusion of all 'pericentral' cells in one category would yield little overall difference from ICC, as in Müller-Preuss et al. (1994). However, the contingent of ~20–30% of low-pass (1–2 Hz BMFs) cells would be found primarily in ICD. This hypothesis is depicted in Fig. 15 for ICX and ICD.

This hypothesis also makes sense in light of the forward connectivity of pericentral IC to thalamic subdivisions (Aitkin, 1986; Wenstrup, 2005). Thalamic subdivisions were studied in unanesthetized primate using AM sounds by Preuss and Müller-Preuss (1990). MGBm was found to have a median BMF of 16 Hz, compared to 8 Hz in MGBd and MGBv (few additional details were given). The tuning to faster AM rates in MGBm is consistent with the continuation of the periodotopy of Baumann et al. (2011) into pericentral regions, given the strong input to MGBm from ICX (we emphasize again that this is our own extrapolation and not given directly in their data, despite the high-resolution fMRI used). The distribution of BMFs collapsed across all MGB subdivisions (Müller-Preuss et al., 1988; Preuss and Müller-Preuss, 1990) exhibits, in addition to a large number of cells with 16–64 Hz BMFs, a second mode centered at 4 Hz (Fig. 12). It is our interpretation that the lower mode at 4 Hz is to be associated most strongly with MGBd, and the faster mode with MGBm and MGBv. There are, however, a number of sub-divisions within MGBd, and these are expected to be diverse in their temporal properties, but we do not cover these further distinctions here.

Returning to the question of the origins of fluctuation range tuning in auditory cortex, we can now see two hypotheses which are not mutually exclusive. First, the 'belt' regions of IC and thalamus (mainly ICD, MGBd) include a significant contingent of cells with low BMF tunings in the fluctuation range or simply low-pass. By the first hypothesis, these tunings are fed forward with little additional contribution by the cortex. The second hypothesis is that the belt/parabelt regions of auditory cortex obtain their fluctuation range tuning by their intrinsic cellular and network properties. The most likely overall interpretation is that fluctuation range and low-pass tuning begins to emerge in the pericentral IC and non-lemniscal thalamus, but the final psychophysical characteristic (~2–5 Hz peak) is mostly due to inherent properties of the belt/parabelt cortical regions. That the belt/parabelt cortical regions must themselves play a strong role in their AM tuning properties is supported by their diverse inputs, not only from non-lemniscal regions of thalamus, but also from core thalamus and cortex (MGBv, AI, R). Since these are the well-studied regions with established AM tunings mostly above the ~2–5 Hz range, then the slower response properties of certain belt/parabelt regions must be due at least in part to their own intrinsic processing.

There are at least 2 additional considerations which support an origin for the psychophysical bandpass characteristic with peak at ~2–5 Hz at the cortical or thalamocortical level. First, the same general phenomenon is observed in vision (Bartley, 1939; Fox and Raichle, 1984), except that the peak brightness and visual cortical activation is at ~7 Hz (i.e., just below the α range, whereas the auditory phenomenon is just below the θ range, the respective spontaneous rhythms). Visual information does not pass through extensive brainstem processing, and the phenomenon almost certainly arises at the thalamocortical level. Second, the separability result (Chi et al., 1999; Langers et al., 2003) is compatible with all preprocessing channels converging in the final step on a common bandpass tuning. Thus, the peak at ~2–5 Hz tuning applies to all modulated sounds (AM, FM, noise, tones, ripple sounds) and second-envelope modulation of periodicity pitch AM. The most parsimonious explanation is that all channels must converge to primary and non-primary cortical regions and their intrinsic tuning characteristics.

Finally, we are able to identify a simple and plausible mechanism from more intensive physiological studies of primary auditory cortex, namely an inhibition of some ~25–250 ms duration following the initial excitation (de Ribaupierre et al., 1972; Volkov and Galazyuk, 1991; Depireux et al., 2001; Ojima and Murakami, 2002; Tan et al., 2004; Chang et al., 2005; Sadagopan and Wang, 2010). This has been found with extracellular, intracellular, whole-cell, and *in vitro* recordings, and up to ~50–100 ms this involves an inhibitory (GABAergic) input to the cortical cell, whereas synaptic depression is implicated during the more prolonged phase of the inhibition (Wehr and Zador, 2005). Thus, the intrinsic inhibitory circuitry (along with synaptic depression) exerts a *temporal contrast* upon cortical inputs, with a time course appropriate for the ~2–32 Hz AM tunings found in AI (note that anesthesia may tend to prolong inhibition). For the non-primary cortex, we hypothesize that the same mechanism could play a role in the ~1–10 Hz bandpass tuning to AM, but with a prolonged inhibitory phase to give the slower tuning. If this prolonged inhibitory phase involved greater synaptic depression, this would also be compatible with the general preference for novelty observed in non-primary cortices. Cells exhibiting this temporal contrast (by whatever mechanism) will exhibit 'phasic' response properties, at the appropriate time scale, which could help explain certain 'phasic' results with human fMRI (Giraud et al., 2000; Seifritz et al., 2002).

We therefore conclude that the observed psychophysical tuning to the fluctuation range, with broad peak at ~2–5 Hz, is primarily a function of belt and parabelt regions of the thalamocortical system. Some neurons of the core (MGBv, AI/R) also exhibit tuning in the fluctuation range, so the core regions are not to be entirely excluded from the result. But the belt and parabelt regions represent the

largest territory of the auditory cortex in humans, so their dominance in the psychophysical result is expected on these grounds as well. Moreover, belt and parabelt regions are interconnected with frontal and parietal regions (Jones, 2010) involved in response selection, language (for reporting), and other aspects of conscious behavior (e.g., Romo and de Lafuente, 2013), so this further implicates the belt and parabelt regions in the final psychophysical result. This is also consistent with the association of speech processing at syllabic and word time scales with multi-modal processing, attention, linguistic context, and top-down influences generally.

6. Conclusions

6.1. Signal processing significance

We conclude by briefly considering the *signal processing significance* of the observed fluctuation range (~1–10 Hz) tuning. In speech processing for ASR, the task of separating syllabic or phonemic units from the continuous speech stream is known as *automatic segmentation*, and usually relies on measures of spectral *change* (Sakai and Doshita, 1963; Tappert, 1972) or AM maxima/minima (Mermelstein, 1975; Reddy, 1976; Zwicker et al., 1979). Note that these measures are applied to the output of the auditory periphery model (or critical-band filter bank). The importance of the syllabic unit for ASR in general was advocated in two outstanding publications of the 1970s (Fujimura, 1975; Ruske and Schotola, 1978), and has since been adopted by other ASR workers. However, it was not until Hirsch and colleagues (Hirsch, 1988; Hirsch et al., 1991) that we find filtering in the modulation domain to enhance the speech-related fluctuations for ASR purposes. These studies showed that high-pass filtering of each subband envelope at ~2 Hz improved ASR performance under noisy (reverberant) conditions. Some early users of the cepstrum for automatic speaker verification (Atal, 1974; Furui, 1981) had noted improvements by removing a running average from the cepstral coefficients (effectively a high-pass filter).

Hermansky and co-workers first employed bandpass filtering in the fluctuation range for improving ASR (Hermansky et al., 1991; Hermansky and Morgan, 1994). Relative immunity to steady background noise was achieved in their 'RASTA' system by bandpass filtering the log-envelopes from ~0.26 to 12.8 Hz: "The key idea here is to suppress constant factors in each spectral component of the short-term auditory-like spectrum…" (Hermansky et al., 1991). With Arai and colleagues (Arai et al., 1996, 1999), this idea was extended to the bandpass filtering of cepstral coefficients (a common representation for ASR) and human perceptual experiments. Similar to the Drullman et al. (1994a,b) (Section 2), they found that modulation frequencies in the range 1–16 Hz were most critical for human speech perception. Kanedera et al. (1998) found that the same range was most critical for ASR performance: "most of the useful linguistic information is in modulation frequency components from the range between 1 and 16 Hz, with the dominant component at around 4 Hz."

Following these initial studies, Greenberg and colleagues have been the major proponents of syllable-range processing for ASR (Greenberg, 1997). Thus, the 'modulation spectrogram' of Greenberg and Kingsbury (Greenberg and Kingsbury, 1997; Kingsbury et al., 1998) performed critical-band filtering of noisy speech, followed by bandpass filtering of each subband envelope at ~4 Hz (10 dB down at 0 and 8 Hz): "The emphasis of modulations in the range of 0–8 Hz with peak sensitivity at 4 Hz acts as a matched filter that passes only signals with temporal dynamics characteristic of speech." Wu et al. (1998a,b) introduced a syllable-based ASR system which used 2–8 Hz bandpass filtering of the subband envelopes. Ongoing work from Greenberg continues to emphasize the syllable and syllable-range modulations (Greenberg, 2006; Ghitza and Greenberg, 2009).

Other recent studies have adopted syllable-oriented and/or modulation-filtering approaches for ASR, but we do not survey further because the basic signal processing significance is already clear from these initial studies. Thus, once in the envelope processing domain (i.e., after the auditory periphery), a matched filter to the long-term envelope spectrum of speech (see Fig. 5) would be tuned to ~1–10 Hz modulations (peak at ~2–5 Hz). Even for the processing of natural sounds, events (or, broadly speaking, sounds that come and go) are of more ethological significance compared to steady background sounds, so the high-pass portion of the modulation tuning curve makes sense for general mammalian auditory processing as well. All of the ASR processing schemes mentioned above employ the fluctuation range modulation filtering as the final stage before recognition (in any case, after the ~cochlear filter-bank and other transformations such as extracting cepstral coefficients). This matches the overall model for mammalian auditory processing (Fig. 15), where it is only in the final stages, perhaps not until belt/parabelt auditory cortices, that the fluctuation range modulation tuning emerges. Thus, the results of auditory peripheral and brainstem processing (pitch, rapid onsets, etc.) are submitted to a final temporal contrast with excitatory/inhibitory phases of appropriate duration to yield the ~1–10 Hz tuning, prior to recognition.

6.2. Summary

The most useful contribution of this review is to gather together in one place the studies from psychophysics and neurophysiology concerning the fluctuation range (~1–10 Hz) of modulated sounds. The relevance to the speech syllabic rate (~2–5 Hz) was discussed throughout. We recovered the pre-1970s finding that human sensitivity to AM and FM sounds exhibits a bandpass characteristic with a peak at ~2–5 Hz, and found that the fMRI and animal neurophysiology evidence is so far consistent with this bandpass characteristic. But this is only the starting point; particularly for physiological studies of non-core regions, we were forced to make use of extremely limited existing data. The present survey clearly indicates that more complete evidence remains to be desired from animal and human neurophysiology. The present review highlights the need, with respect to speech, to include both of the "two systems" (Andersen et al., 1980; Aitkin, 1986) in such models. That is, the "second" and most-often neglected system, consisting of non-core divisions of the IC, MGB, and auditory cortex, appear to be critical for ~1–10 Hz AM/FM processing and therefore speech perception.

Acknowledgments

We thank Brian Malone for helpful discussion and comments on the manuscript, and Dora Hermes for helpful discussion at early stages. This work was supported by NINDS fellowship F32-NS061616 (EE) and NIH grants R00-NS065120, R01-DC012379, DP2 OD008627 (EC).

References

Ades, H.W., 1959. Central auditory mechanisms. In: Magoun, H.W. (Ed.), Handbook of Physiology, Section 1, vol. I. American Physiological Society, Washington, DC, pp. 585–613.
Aiken, S.J., Picton, T.W., 2008. Human cortical responses to the speech envelope. Ear Hear. 29, 139–157.
Aitkin, L.M., 1986. The Auditory Midbrain: Structure and Function in the Central Auditory Pathway. Humana Press, Clifton, NJ.
Aitkin, L.M., Merzenich, M.M., Irvine, D.R.F., Clarey, J.C., Nelson, J.E., 1986. Frequency representation in auditory cortex of the common marmoset (*Callithrix jacchus jacchus*). J. Comp. Neurol. 252, 175–185.

Andersen, R.A., Knight, P.L., Merzenich, M.M., 1980. The thalamocortical and corticothalamic connections of AI, AII, and the anterior auditory field (AAF) in the cat: evidence for two largely segregated systems of connections. J. Comp. Neurol. 194, 663–701.
Arai, T., Pavel, M., Hermansky, H., Avendano, C., 1996. Intelligibility of speech with filtered time trajectories of spectral envelopes. In: Proceedings of the International Conference on Spoken Language (ICSLP), Philadelphia, PA, vol. 4. IEEE, pp. 2490–2493.
Arai, T., Pavel, M., Hermansky, H., Avendano, C., 1999. Syllable intelligibility for temporally filtered LPC cepstral trajectories. J. Acoust. Soc. Am. 105, 2783–2791.
Atal, B.S., 1974. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. J. Acoust. Soc. Am. 55, 1304–1322.
Atlas, L.E., Shamma, S.A., 2003. Joint acoustic and modulation frequency. EURASIP J. Adv. Signal Process. 2003, 668–675.
Bacon, S.P., Viemeister, N.F., 1985. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. Audiology 24, 117–134.
Bartley, S.H., 1939. Some factors in brightness discrimination. Psychol. Rev. 46, 337–358.
Baumann, S., Griffiths, T.D., Sun, L., Petkov, C.I., Thiele, A., Rees, A., 2011. Orthogonal representation of sound dimensions in the primate midbrain. Nat. Neurosci. 14, 423–425.
Baumann, S., Petkov, C.I., Griffiths, T.D., 2013. A unified framework for the organization of the primate auditory cortex. Front. Syst. Neurosci. 7, 1–8.
Bendor, D., Wang, X., 2008. Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. J. Neurophysiol. 100, 888–906.
Berger, H., 1930. Über das Elektrenkephalogramm des Menschen. II. J. Psychol. Neurol. 40, 160–179.
Bieser, A., Müller-Preuss, P., 1996. Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp. Brain Res. 108, 273–284.
Bilsen, F.A., Ritsma, R.J., 1967. Repetition pitch mediated by temporal fine structure at dominant spectral regions. Acustica 19, 114–116.
Bilsen, F.A., Wieman, J.L., 1980. Atonal periodicity sensation for comb filtered noise signals. In: van den Brink, G., Bilsen, F.A. (Eds.), Psychophysical, Physiological, and Behavioural Studies in Hearing. Delft Univ. Press, Delft, The Netherlands, pp. 379–383.
Binder, J.R., Rao, S.M., Hammeke, T.A., Frost, J.A., Bandettini, P.A., Hyde, J.S., 1994a. Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. Brain Res. Cogn. Brain Res. 2, 31–38.
Binder, J.R., Rao, S.M., Hammeke, T.A., Yetkin, F.Z., Jesmanowicz, A., Bandettini, P.A., Wong, E.C., Estkowski, L.D., Goldstein, M.D., Haughton, V.M., Hyde, J.S., 1994b. Functional magnetic resonance imaging of human auditory cortex. Ann. Neurol. 35, 662–672.
Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci. 8, 389–395.
Bregman, A.S., 1990. Auditory Scene Analysis: the Perceptual Organization of Sound. MIT Press, Cambridge, MA.
Bremer, F., 1952. Analyse oscillographique des réponses sensorielles des écorces cérébrales et cérébelleuse. Rev. Neurol. 87, 65–92.
Brugge, J.F., Volkov, I.O., Oya, H., Kawasaki, H., Reale, R.A., Fenoy, A.J., Steinschneider, M., Howard III, M.A., 2008. Functional localization of auditory cortical fields of human: click-train stimulation. Hear. Res. 238, 12–24.
Burns, E.M., Viemeister, N.F., 1976. Nonspectral pitch. J. Acoust. Soc. Am. 60, 863–869.
Burns, E.M., Viemeister, N.F., 1981. Played-again SAM: further observations on the pitch of amplitude-modulated noise. J. Acoust. Soc. Am. 70, 1655–1660.
Calford, M.B., Aitkin, L.M., 1983. Ascending projections to the medial geniculate body of the cat: evidence for multiple, parallel auditory pathways through thalamus. J. Neurosci. 3, 2365–2380.
Chang, E.F., Bao, S., Imaizumi, K., Schreiner, C.E., Merzenich, M.M., 2005. Development of spectral and temporal response selectivity in the auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 102, 16460–16465.
Chang, H.-T., 1960. Some observations on the excitability changes of cortical and subcortical neurons and their possible significance in the process of conditioning. Electroencephalogr. Clin. Neurophysiol. Suppl. 13, 39–49.
Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.
Chi, T.-S., Gao, Y., Guyton, M.C., Ru, P., Shamma, S.A., 1999. Spectro-temporal modulation transfer functions and speech intelligibility. J. Acoust. Soc. Am. 106, 2719–2732.
Clare, M.H., Bishop, G.H., 1952. The intracortical excitability cycle following stimulation of the optic pathway of the cat. Electroencephalogr. Clin. Neurophysiol. 4, 311–320.
Clark, P., Atlas, L.E., 2009. Time-frequency coherent modulation filtering of nonstationary signals. IEEE Trans. Signal Process. 57, 4323–4332.
Coleman, R.F., 1971. Effect of waveform changes upon roughness perception. Folia Phoniatr. 23, 314–322.
Crone, N.E., Boatman, D., Gordon, B., Hao, L., 2001. Induced electrocorticographic gamma activity during auditory perception. Clin. Neurophysiol. 112, 565–582.
Crone, N.E., Miglioretti, D.L., Gordon, B., Lesser, R.P., 1998. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain 121, 2301–2315.
Curio, G., 2000. Linking 600-Hz "spikelike" EEG/MEG wavelets ("sigma-bursts") to cellular substrates: concepts and caveats. J. Clin. Neurophysiol. 17, 377–396.
Darrow, C.W., Graf, C.G., 1945. Relation of electroencephalogram to photometrically observed vasomotor changes in the brain. J. Neurophysiol. 8, 449–461.
Dau, T., Kollmeier, B., Kohlrausch, A., 1997. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892–2905.
Davis, H., Davis, P.A., Loomis, A.L., Harvey, E.N., Hobart III, G.A., 1937. Changes in human brain potentials during the onset of sleep. Science 86, 448–450.
de Ribaupierre, F., Goldstein Jr., M.H., Yeni-Komshian, G.H., 1972. Intracellular study of the cat's primary auditory cortex. Brain Res. 48, 185–204.
Depireux, D.A., Simon, J.Z., Klein, D.J., Shamma, S.A., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. J. Neurophysiol. 85, 1220–1234.
Drullman, R., Festen, J.M., Plomp, R., 1994a. Effect of reducing slow temporal modulations on speech reception. J. Acoust. Soc. Am. 95, 2670–2680.
Drullman, R., Festen, J.M., Plomp, R., 1994b. Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Am. 95, 1053–1064.
Dubrovskii, N.A., Tumarkina, L.N., 1967. Investigations of the human perception of amplitude-modulated noise. Sov. Phys. Acoust. 13, 41–47.
Ectors, L., 1936. Étude de l'activité électrique du cortex cérébral chez le lapin non narcotisé ni curarisé. Arch. Int. Physiol. 43, 267–298.
Eggermont, J.J., 1993. Differential effects of age on click-rate and amplitude modulation-frequency coding in primary auditory cortex of the cat. Hear. Res. 65, 175–192.
Eggermont, J.J., 1994. Temporal modulation transfer functions for AM and FM stimuli in cat auditory cortex. Effects of carrier type, modulating waveform and intensity. Hear. Res. 74, 51–66.
Eggermont, J.J., 1998. Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J. Neurophysiol. 80, 2743–2764.
Eggermont, J.J., 2002. Temporal modulation transfer functions in cat primary auditory cortex: separating stimulus effects from neural mechanisms. J. Neurophysiol. 87, 305–321.
Elliott, T.M., Theunissen, F.E., 2009. The modulation transfer function for speech intelligibility. PLoS Comput. Biol. 5, e1000302.
Erulkar, S.D., Rose, J.E., Davies, P.W., 1956. Single unit activity in the auditory cortex of the cat. Bull. Johns Hopkins Hosp. 99, 55–86.
Evans, E.F., 1968. Upper and lower levels of the auditory system: a contrast of structure and function. In: Caianiello, E.R. (Ed.), Neural Networks. Springer-Verlag, Berlin, New York, pp. 24–33.
Fastl, H., 1977. Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise. In: Evans, E.F., Wilson, J.P. (Eds.), Psychophysics and Physiology of Hearing. Academic Press, London, New York, pp. 403–417.
Fastl, H., 1982. Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise. Hear. Res. 8, 59–69.
Fastl, H., 1983. Fluctuation strength of modulated tones and broadband noise. In: Klinke, R., Hartmann, R. (Eds.), Hearing, Physiological Bases and Psychophysics. Springer-Verlag, Berlin, New York, pp. 282–286.
Fastl, H., Hesse, A., Schorer, E., Urbas, J.V., Müller-Preuss, P., 1986. Searching for neural correlates of the hearing sensation fluctuation strength in the auditory cortex of squirrel monkeys. Hear. Res. 23, 199–203.
Fastl, H., Schorer, E., 1986. Critical bandwidth at low frequencies reconsidered. In: Moore, B.C.J., Patterson, R.D. (Eds.), Auditory Frequency Selectivity. Plenum Press, New York, pp. 311–322.
Fastl, H., Stoll, G., 1979. Scaling of pitch strength. Hear. Res. 1, 293–301.
Fastl, H., Zwicker, E., 2007. Psychoacoustics: Facts and Models. Springer, Berlin, New York.
Flanagan, J.L., 1961. Audibility of periodic pulses and a model for the threshold. J. Acoust. Soc. Am. 33, 1540–1549.
Fox, P.T., Raichle, M.E., 1984. Stimulus rate dependence of regional cerebral blood flow in human striate cortex, demonstrated by positron emission tomography. J. Neurophysiol. 51, 1109–1120.
Frith, C.D., Friston, K.J., 1996. The role of the thalamus in "top down" modulation of attention to sound. Neuroimage 4, 210–215.
Fujimura, O., 1975. Syllable as a unit of speech recognition. IEEE Trans. Acoust. Speech Signal Process. 23, 82–87.
Fullerton, B.C., Pandya, D.N., 2007. Architectonic analysis of the auditory-related areas of the superior temporal region in human brain. J. Comp. Neurol. 504, 470–498.
Furui, S., 1981. Cepstral analysis technique for automatic speaker verification. IEEE Trans. Acoust. Speech Signal Process. 29, 254–272.
Galambos, R., 1960. Studies of the auditory system with implanted electrodes. In: Rasmussen, G.L., Windle, W.F. (Eds.), Neural Mechanisms of the Auditory and Vestibular Systems. C.C. Thomas, Springfield, IL, pp. 137–151.
Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. Proc. Natl. Acad. Sci. U. S. A. 78, 2643–2647.
Ghitza, O., Greenberg, S., 2009. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica 66, 113–126.
Giraud, A.-L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I.S., Frackowiak, R.S.J., Kleinschmidt, A., 2000. Representation of the temporal envelope of sounds in the human brain. J. Neurophysiol. 84, 1588–1598.
Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517.

Goldstein Jr., M.H., Kiang, N.Y.-S., Brown, R.M., 1959. Responses of the auditory cortex to repetitive acoustic stimuli. J. Acoust. Soc. Am. 31, 356–364.
Goldstein Jr., M.H., Knight, P.L., 1980. Comparative organization of mammalian auditory cortex. In: Popper, A.N., Fay, R.R. (Eds.), Comparative Studies of Hearing in Vertebrates. Springer-Verlag, New York, pp. 375–398.
Grass, A.M., Gibbs, F.A., 1938. A Fourier transform of the electroencephalogram. J. Neurophysiol. 1, 521–526.
Green, D.M., 1973. Minimum integration time. In: Møller, A.R. (Ed.), Basic Mechanisms in Hearing. Academic Press, New York, London, pp. 829–846.
Greenberg, S., 1997. On the origins of speech intelligibility in the real world. In: Proceedings of the ESCA Workshop on Robust Speech Recognition for Unknown Communication Channels, pp. 23–32.
Greenberg, S., 2006. A multi-tier framework for understanding spoken language. In: Greenberg, S., Ainsworth, W.A. (Eds.), Listening to Speech: an Auditory Perspective. Lawrence Erlbaum Assoc., Mahwah, NJ, pp. 411–433.
Greenberg, S., Kingsbury, B.E.D., 1997. The modulation spectrogram: in pursuit of an invariant representation of speech. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP). IEEE, vol. 3, pp. 1647–1650.
Hackett, T.A., Preuss, T.M., Kaas, J.H., 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J. Comp. Neurol. 441, 197–222.
Hall, D.A., 2005. Sensitivity to spectral and temporal properties of sound in human non-primary auditory cortex. In: König, R., et al. (Eds.), The Auditory Cortex: a Synthesis of Human and Animal Research. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 51–76.
Hall, D.A., 2012. fMRI of the central auditory system. In: Faro, S.H., Mohamed, F.B. (Eds.), Functional Neuroradiology: Principles and Clinical Applications. Springer, New York, pp. 575–591.
Hall, D.A., Johnsrude, I.S., Haggard, M.P., Palmer, A.R., Akeroyd, M.A., Summerfield, A.Q., 2002. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 12, 140–149.
Harms, M.P., Melcher, J.R., 2002. Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. J. Neurophysiol. 88, 1433–1450.
Hart, H.C., Palmer, A.R., Hall, D.A., 2003. Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. Cereb. Cortex 13, 773–781.
Helmholtz, H.v., 1863. Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. F. Vieweg u. Sohn, Braunschweig.
Hermansky, H., Morgan, N., 1994. RASTA processing of speech. IEEE Trans. Speech Audio Process. 2, 578–589.
Hermansky, H., Morgan, N., Bayya, A., Kohn, P., 1991. The challenge of inverse-E: the RASTA-PLP method. In: Proceedings of the Asilomar Conference on Signals, Systems and Computers, IEEE, vol. 2, pp. 800–804.
Hirsch, H.-G., 1988. Automatic speech recognition in rooms. In: Proceedings of the European Signal Processing Conference (EUSIPCO), North-Holland, vol. 3, pp. 1177–1180.
Hirsch, H.-G., Meyer, P., Ruehl, H.-W., 1991. Improved speech recognition using high-pass filtering of subband envelopes. In: Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH). ISCA, pp. 413–416.
Houtgast, T., Steeneken, H.J.M., 1973. The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 28, 66–73.
Houtgast, T., Steeneken, H.J.M., Plomp, R., 1980. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acustica 46, 60–72.
Imaizumi, K., Lee, C.C., Linden, J.F., Winer, J.A., Schreiner, C.E., 2005. The anterior field of auditory cortex: neurophysiological and neuroanatomical organization. In: König, R., et al. (Eds.), The Auditory Cortex: a Synthesis of Human and Animal Research. Lawrence Erlbaum Assoc., Mahwah, NJ, pp. 95–110.
Imig, T.J., Ruggero, M.A., Kitzes, L.M., Javel, E., Brugge, J.F., 1977. Organization of auditory cortex in the owl monkey (*Aotus trivirgatus*). J. Comp. Neurol. 171, 111–128.
Irvine, D.R.F., 1986. The Auditory Brainstem: a Review of the Structure and Function of Auditory Brainstem Processing Mechanisms. Springer-Verlag, Berlin, Heidelberg.
Jasper, H.H., Andrews, H.L., 1938. Electroencephalography: III. Normal differentiation between occipital and pre-central regions in man. Arch. Neurol. Psychiatr. 39, 96–115.
Jones, E.G., 2010. The historical development of ideas about the auditory cortex. In: Winer, J.A., Schreiner, C.E. (Eds.), The Auditory Cortex. Springer, New York, pp. 1–40.
Joris, P.X., Schreiner, C.E., Rees, A., 2004. Neural processing of amplitude-modulated sounds. Physiol. Rev. 84, 541–577.
Jung, R., 1941. Das Elektrencephalogramm und seine klinische Anwendung. II. Das EEG des Gesunden, seine Variationen und Veränderungen und deren Bedeutung für das pathologische EEG. Nervenarzt 14 (57–70), 104–117.
Kaas, J.H., Hackett, T.A., 2000. Subdivisions of auditory cortex and processing streams in primates. Proc. Natl. Acad. Sci. U. S. A. 97, 11793–11799.
Kaas, J.H., Hackett, T.A., Tramo, M.J., 1999. Auditory processing in primate cerebral cortex. Curr. Opin. Neurobiol. 9, 164–170.
Kanedera, N., Hermansky, H., Arai, T., 1998. On properties of modulation spectrum for robust automatic speech recognition. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), IEEE, vol. 2, pp. 613–616.
Katsuki, Y., Murata, K., Suga, N., Takenaka, T., 1960. Single unit activity in the auditory cortex of an unanaesthetized monkey. Proc. Jpn. Acad. 36, 435–438.
Kay, R.H., 1982. Hearing of modulation in sounds. Physiol. Rev. 62, 894–975.
Kay, R.H., Matthews, D.R., 1972. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. J. Physiol. 225, 657–677.
Kemp, S., 1982. Roughness of frequency modulated tones. Acustica 50, 126–133.
Kenmochi, M., Eggermont, J.J., 1997. Autonomous cortical rhythms affect temporal modulation transfer functions. Neuroreport 8, 1589–1593.
Kingsbury, B.E.D., Morgan, N., Greenberg, S., 1998. Robust speech recognition using the modulation spectrogram. Speech Commun. 25, 117–132.
Langers, D.R.M., Backes, W.H., van Dijk, P., 2003. Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. Neuroimage 20, 265–275.
Langner, G., 1992. Periodicity coding in the auditory system. Hear. Res. 60, 115–142.
Liang, L., Lu, T., Wang, X., 2002. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J. Neurophysiol. 87, 2237–2261.
Licklider, J.C.R., 1959. Three auditory theories. In: Koch, S. (Ed.), Psychology: a Study of a Science, vol. 1. McGraw-Hill, New York, pp. 41–144.
Ljung, L., 1999. System Identification: Theory for the User. Prentice Hall PTR, Upper Saddle River, NJ.
Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157.
Malone, B.J., Schreiner, C.E., 2010. Time-varying sounds: amplitude envelope modulations. In: Rees, A., Palmer, A.R. (Eds.), The Auditory Brain. Oxford Univ. Press, Oxford, New York, pp. 125–148.
Malone, B.J., Scott, B.H., Semple, M.N., 2007. Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. J. Neurophysiol. 98, 1451–1474.
Mermelstein, P., 1975. Automatic segmentation of speech into syllabic units. J. Acoust. Soc. Am. 58, 880–883.
Merzenich, M.M., Brugge, J.F., 1973. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res. 50, 275–296.
Møller, A.R., 1972a. Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat. Acta Physiol. Scand. 86, 223–238.
Møller, A.R., 1972b. Coding of sounds in lower levels of the auditory system. Q. Rev. Biophys. 5, 59–155.
Morest, D.K., Oliver, D.L., 1984. The neuronal architecture of the inferior colliculus in the cat: defining the functional anatomy of the auditory midbrain. J. Comp. Neurol. 222, 209–236.
Mukamel, R., Gelbard, H., Arieli, A., Hasson, U., Fried, I., Malach, R., 2005. Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. Science 309, 951–954.
Müller-Preuss, P., Bieser, A., Preuss, A., Fastl, H., 1988. Neural processing of AM-sounds within central auditory pathway. In: Syka, J., Masterton, R.B. (Eds.), Auditory Pathway: Structure and Function. Plenum Press, New York, pp. 327–331.
Müller-Preuss, P., Flachskamm, C., Bieser, A., 1994. Neural encoding of amplitude modulation within the auditory midbrain of squirrel monkeys. Hear. Res. 80, 197–208.
Nourski, K.V., Brugge, J.F., 2011. Representation of temporal sound features in the human auditory cortex. Rev. Neurosci. 22, 187–203.
Obleser, J., Herrmann, B., Henry, M.J., 2012. Neural oscillations in speech: don't be enslaved by the envelope. Front. Hum. Neurosci. 6, 1–4.
O'Connor, K.N., Johnson, J.S., Niwa, M., Noriega, N.C., Marshall, E.A., Sutter, M.L., 2011. Amplitude modulation detection as a function of modulation frequency and stimulus duration: comparisons between macaques and humans. Hear. Res. 277, 37–43.
Ohala, J.J., 1975. The temporal regulation of speech. In: Fant, G., Tatham, M.A.A. (Eds.), Auditory Analysis and Perception of Speech. Academic Press, London, New York, pp. 431–453.
Ojima, H., Murakami, K., 2002. Intracellular characterization of suppressive responses in supragranular pyramidal neurons of cat primary auditory cortex *in vivo*. Cereb. Cortex 12, 1079–1091.
Oliver, D.L., 2005. Neuronal organization in the inferior colliculus. In: Winer, J.A., Schreiner, C.E. (Eds.), The Inferior Colliculus. Springer, New York, pp. 69–114.
Patterson, R.D., Johnson-Davies, D., Milroy, R., 1978. Amplitude-modulated noise: the detection of modulation versus the detection of modulation rate. J. Acoust. Soc. Am. 63, 1904–1911.
Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3, 1–17.
Pfurtscheller, G., 1999. EEG event-related desynchronization (ERD) and event-related synchronization (ERS). In: Niedermeyer, E., Lopes da Silva, F. (Eds.), Electroencephalography: Basic Principles, Clinical Applications, and Related Fields. Williams & Wilkins, Baltimore, London, pp. 958–967.
Picinbono, B., 1997. On instantaneous amplitude and phase of signals. IEEE Trans. Signal Process. 45, 552–560.
Picton, T.W., John, M.S., Dimitrijevic, A., Purcell, D.W., 2003. Human auditory steady-state responses. Int. J. Audiol. 42, 177–219.
Picton, T.W., Skinner, C.R., Champagne, S.C., Kellett, A.J., Maiste, A.C., 1987. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J. Acoust. Soc. Am. 82, 165–178.
Plomp, R., Houtgast, T., Steeneken, H.J.M., 1984. The modulation transfer function in audition. In: van Doorn, A.J., et al. (Eds.), Limits in Perception: Essays in Honour of Maarten A. Bouman. VNU Science Press, Utrecht, pp. 117–138.

Plomp, R., Steeneken, H.J.M., 1968. Interference between two simple tones. J. Acoust. Soc. Am. 43, 883–884.
Pollack, I., 1951. On the threshold of loudness of repeated bursts of noise. J. Acoust. Soc. Am. 23, 646–650.
Potter, R.K., Kopp, G.A., Green, H.C., 1947. Visible Speech. D. Van Nostrand, New York.
Preuss, A., Müller-Preuss, P., 1990. Processing of amplitude modulated sounds in the medial geniculate body of squirrel monkeys. Exp. Brain Res. 79, 207–211.
Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97, 11800–11806.
Reddy, D.R., 1976. Speech recognition by machine: a review. Proc IEEE 64, 501–531.
Riesz, R.R., 1928. Differential intensity sensitivity of the ear for pure tones. Phys. Rev. 31, 867–875.
Rinne, T., Pekkola, J., Degerman, A., Autti, T., Jääskeläinen, I.P., Sams, M., Alho, K., 2005. Modulation of auditory cortex activation by sound presentation rate and attention. Hum. Brain Mapp. 26, 94–99.
Ritsma, R.J., 1962. Existence region of the tonal residue. I. J. Acoust. Soc. Am. 34, 1224–1229.
Rodenburg, M., 1972. Sensitivity of the auditory system to differences in intensity, Ph.D. Thesis, Rotterdam, Medical Faculty, Erasmus University Rotterdam.
Rodenburg, M., 1977. Investigation of temporal effects with amplitude modulated signals. In: Evans, E.F., Wilson, J.P. (Eds.), Psychophysics and Physiology of Hearing: an International Symposium. Academic Press, London, New York, pp. 429–439.
Romo, R., de Lafuente, V., 2013. Conversion of sensory signals into perceptual decisions. Prog. Neurobiol. 103, 41–75.
Rose, J.E., Woolsey, C.N., 1949. The relations of thalamic connections, cellular structure and evocable electrical activity in the auditory region of the cat. J. Comp. Neurol. 91, 441–466.
Rose, J.E., Woolsey, C.N., 1958. Cortical connections and functional organization of the thalamic auditory system of the cat. In: Harlow, H.F., Woolsey, C.N. (Eds.), Biological and Biochemical Bases of Behavior. Univ. of Wisconsin Press, Madison, pp. 127–150.
Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B Biol. Sci. 336, 367–373.
Ruske, G., Schotola, T., 1978. An approach to speech recognition using syllabic decision units. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 3. IEEE, pp. 722–725.
Sadagopan, S., Wang, X., 2010. Contribution of inhibition to stimulus selectivity in primary auditory cortex of awake primates. J. Neurosci. 30, 7314–7325.
Sakai, T., Doshita, S., 1963. The automatic speech recognition system for conversational sound. IEEE Trans. Electron Comput. 12, 835–846.
Scheeringa, R., Petersson, K.M., Oostenveld, R., Norris, D.G., Hagoort, P., Bastiaansen, M.C.M., 2009. Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. Neuroimage 44, 1224–1238.
Schimmel, O., van de Par, S., Breebaart, J., Kohlrausch, A., 2008. Sound segregation based on temporal envelope structure and binaural cues. J. Acoust. Soc. Am. 124, 1130–1145.
Schönwiesner, M., Zatorre, R.J., 2009. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. U. S. A. 106, 14611–14616.
Schreiner, C.E., Langner, G., 1988. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J. Neurophysiol. 60, 1823–1840.
Schreiner, C.E., Urbas, J.V., 1988. Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. Hear. Res. 32, 49–63.
Scott, B.H., Malone, B.J., Semple, M.N., 2011. Transformation of temporal processing across auditory cortex of awake macaques. J. Neurophysiol. 105, 712–730.
Scott, S.K., Rosen, S., Lang, H., Wise, R.J.S., 2006. Neural correlates of intelligibility in speech investigated with noise vocoded speech – a positron emission tomography study. J. Acoust. Soc. Am. 120, 1075–1083.
Seashore, C.E., 1936. Psychology of the Vibrato in Voice and Instrument. The University Press, Iowa City.
Seifritz, E., Di Salle, F., Esposito, F., Bilecen, D., Neuhoff, J.G., Scheffler, K., 2003. Sustained blood oxygenation and volume response to repetition rate-modulated sound in human auditory cortex. Neuroimage 20, 1365–1370.
Seifritz, E., Esposito, F., Hennel, F., Mustovic, H., Neuhoff, J.G., Bilecen, D., Tedeschi, G., Scheffler, K., Di Salle, F., 2002. Spatiotemporal pattern of neural processing in the human auditory cortex. Science 297, 1706–1708.
Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Sheer, D.E., 1989. Sensory and cognitive 40-Hz event-related potentials: behavioral correlates, brain function, and clinical application. In: Basar, E., Bullock, T.H. (Eds.), Brain Dynamics: Progress and Perspectives. Springer-Verlag, Berlin, New York, pp. 339–374.
Shower, E.G., Biddulph, R., 1931. Differential pitch sensitivity of the ear. J. Acoust. Soc. Am. 3, 275–287.
Singh, N.C., Theunissen, F.E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. J. Acoust. Soc. Am. 114, 3394–3411.
Stott, A., Axon, P.E., 1955. The subjective discrimination of pitch and amplitude fluctuations in recording systems. Proc. IEE B Radio Electron. Eng. 102, 643–656.
Sweet, R.A., Dorph-Petersen, K.-A., Lewis, D.A., 2005. Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J. Comp. Neurol. 491, 270–289.
Tan, A.Y.Y., Zhang, L.I., Merzenich, M.M., Schreiner, C.E., 2004. Tone-evoked excitatory and inhibitory synaptic conductances of primary auditory cortex neurons. J. Neurophysiol. 92, 630–643.
Tanaka, H., Fujita, N., Watanabe, Y., Hirabuki, N., Takanashi, M., Oshiro, Y., Nakamura, H., 2000. Effects of stimulus rate on the auditory cortex using fMRI with 'sparse' temporal sampling. Neuroreport 11, 2045–2049.
Tappert, C.C., 1972. A preliminary investigation of adaptive control in the interaction between segmentation and segment classification in automatic recognition of continuous speech. IEEE Trans. Syst. Man Cybern. 2, 66–72.
Terhardt, E, 1968. Über akustische Rauhigkeit und Schwankungsstärke, 24. Acustica 30, 215–224.
Terhardt, E., 1970. Frequency analysis and periodicity detection in the sensations of roughness and periodicity pitch. In: Plomp, R., Smoorenburg, G.F. (Eds.), Frequency Analysis and Periodicity Detection in Hearing. A. W. Sijthoff, Leiden, pp. 278–287.
Terhardt, E., 1974. On the perception of periodic sound fluctuations (roughness). Acustica 30, 201–213.
Tian, B., Rauschecker, J.P., 2004. Processing of frequency-modulated sounds in the lateral auditory belt cortex of the rhesus monkey. J. Neurophysiol. 92, 2993–3013.
Tonndorf, J., Brogan, F.A., Washburn, D.D., 1955. Auditory D.L. of intensity in normal hearing subjects. AMA Arch. Otolaryngol. 62, 292–305.
van Zanten, G.A., 1980. Temporal modulation transfer functions for intensity modulated noise. In: van den Brink, G., Bilsen, F.A. (Eds.), Psychophysical, Physiological, and Behavioural Studies in Hearing. Delft University Press, Delft, The Netherlands, pp. 206–209.
van Zanten, G.A., Senten, C.J.J., 1983. Spectro-temporal modulation transfer function (STMTF) for various types of temporal modulation and a peak distance of 200 Hz. J. Acoust. Soc. Am. 74, 52–62.
Viemeister, N.F., 1973. Temporal modulation transfer functions for audition (A). J. Acoust. Soc. Am. 53, 312.
Viemeister, N.F., 1977. Temporal factors in audition: a systems analysis approach. In: Evans, E.F., Wilson, J.P. (Eds.), Psychophysics and Physiology of Hearing. Academic Press, London, New York, pp. 419–428.
Viemeister, N.F., 1979. Temporal modulation transfer functions based on modulation thresholds. J. Acoust. Soc. Am. 66, 1364–1380.
Viemeister, N.F., Plack, C.J., 1993. Time analysis. In: Yost, W.A., et al. (Eds.), Human Psychophysics, vol. 3. Springer-Verlag, New York.
Volkov, I.O., Galazyuk, A.V., 1991. Formation of spike response to sound tones in cat auditory cortex neurons: interaction of excitatory and inhibitory effects. Neuroscience 43, 307–321.
Walter, W.G., 1936. The location of cerebral tumors by electro-encephalography. Lancet 228, 305–308.
Walter, W.G., Dovey, V.J., 1944. Electroencephalography in cases of sub-cortical tumor. J. Neurol. Neurosurg. Psychiatry 7, 57–65.
Wang, X., Lu, T., Bendor, D., Bartlett, E.L., 2008. Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 157, 484–494.
Wehr, M., Zador, A.M., 2005. Synaptic mechanisms of forward suppression in rat auditory cortex. Neuron 47, 437–445.
Wendhal, R.W., 1966a. Laryngeal analog synthesis of jitter and shimmer auditory parameters of harshness. Folia Phoniatr. 18, 98–108.
Wendhal, R.W., 1966b. Some parameters of auditory roughness. Folia Phoniatr. 18, 26–32.
Wenstrup, J.J., 2005. The tectothalamic system. In: Winer, J.A., Schreiner, C.E. (Eds.), The Inferior Colliculus. Springer, New York, pp. 200–230.
Wever, E.G., 1929. Beats and related phenomena resulting from the simultaneous sounding of two tones – I. Psychol. Rev. 36, 402–418.
Whitfield, I.C., 1957. The electrical responses of the unanaesthetised auditory cortex in the intact cat. Electroencephalogr. Clin. Neurophysiol. 9, 35–42.
Whitfield, I.C., Evans, E.F., 1965. Responses of auditory cortical neurons to stimuli of changing frequency. J. Neurophysiol. 28, 655–672.
Woolsey, C.N., 1961. Organization of cortical auditory system: review and a synthesis. In: Rosenblith, W.A. (Ed.), Sensory Communication, Contributions. M.I.T. Press, Cambridge, MA, pp. 235–257.
Wu, S.-L., Kingsbury, B.E.D., Morgan, N., Greenberg, S., 1998a. Incorporating information from syllable-length time scales into automatic speech recognition. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 2. IEEE, pp. 721–724.
Wu, S.-L., Kingsbury, B.E.D., Morgan, N., Greenberg, S., 1998b. Performance improvements through combining phone- and syllable-length information in automatic speech recognition. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP). ISCA, pp. 854–857.
Yin, P., Johnson, J.S., O'Connor, K.N., Sutter, M.L., 2011. Coding of amplitude modulation in primary auditory cortex. J. Neurophysiol. 105, 582–600.
Yost, W.A., 1982. The dominance region and ripple noise pitch: a test of the peripheral weighting model. J. Acoust. Soc. Am. 72, 416–425.
Yost, W.A., Hill, R., Perez-Falcon, T., 1978. Pitch and pitch discrimination of broadband signals with rippled power spectra. J. Acoust. Soc. Am. 63, 1166–1175.
Zwicker, E., 1952. Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. Acustica 2, 125–133.
Zwicker, E., Feldtkeller, R., 1967. Das Ohr als Nachrichtenempfänger. Hirzel, Stuttgart.
Zwicker, E., Feldtkeller, R., 1998. The Ear as a Communication Receiver. Acoustical Society of America, Woodbury, NY.
Zwicker, E., Terhardt, E., Paulus, E., 1979. Automatic speech recognition using psychoacoustic models. J. Acoust. Soc. Am. 65, 487–498.