

The effects of temporal asymmetry on the detection and perception of short chirps¹

Stefan Uppenkamp^{a,*}, Sandra Fobel^b, Roy D. Patterson^a

^a Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge CB2 3EG, UK

^b AG Medizinische Physik, Fachbereich Physik, Universität Oldenburg, 26111 Oldenburg, Germany

Received 30 January 2001; accepted 12 April 2001

Abstract

There is an intriguing contrast between the physiological response to short frequency sweeps in the brainstem and the perception produced by these sounds. Dau et al. (2000) demonstrated that optimised chirps with increasing instantaneous frequency (up-chirps), designed to compensate for spatial dispersion along the cochlea, enhance wave V of the auditory brainstem response (ABR), by synchronising excitation of all frequency channels across the basilar membrane. Down-chirps, that is up-chirps reversed in time, increase cochlear phase delays and therefore result in a poor ABR wave V. In this study, a set of psychoacoustical experiments with up-chirps and down-chirps has been performed to investigate how these phase changes affect what we hear. The perceptual contrast is different from what was reported at the brainstem level. It is the down-chirp that sounds more compact, despite the poor synchronisation across channels and phase delays up to 20 ms. The perceived ‘compactness’ of a sound is apparently more determined by the fine structure of excitation within each peripheral channel than by between-channel phase differences. This suggests an additional temporal integration mechanism at a higher stage of auditory processing, which effectively removes phase differences between channels. © 2001 Elsevier Science B.V. All rights reserved.

Key words: Chirp signal; Cochlear phase delay; Temporal integration; Auditory image

1. Introduction

The mechanical properties of the cochlea change progressively from base to apex, mainly because the stiffness of the basilar membrane (BM) decreases, and this produces the well-known frequency-to-place mapping.

Waves propagating along the cochlea exhibit dispersion, that is, their speed is dependent on frequency. As a consequence, the temporal concentration of a very short rectangular pulse (a click) is smeared out in time, with excitation in the lowest frequency channels delayed by about 10 ms with respect to the high-frequency channels. In humans, the dispersion is largely confined to frequencies below 1000 Hz.

Electro-physiological responses to clicks are dominated by the synchronised neural activity in the high-frequency channels. The effect is clear both in the compound action potential (CAP) recorded from the auditory nerve (Johnstone, 1981) and also in the click-evoked, auditory brainstem response (ABR) (Don and Eggermont, 1978). Don et al. (1994) developed a technique of normalising click-evoked ABR by compressing and shifting derived ABR waveforms in the presence of high-pass noise maskers with varying cut-off frequencies. This technique allows one to compensate for individual differences in cochlear response

* Corresponding author. Tel.: +44 (1223) 333-857; Fax: +44 (1223) 333-840.

E-mail address: stefan.uppenkamp@mrc-cbu.cam.ac.uk (S. Uppenkamp).

¹ Parts of this study were presented during the 12th International Symposium on Hearing 2000 in Mierlo/NL (Uppenkamp et al., 2001).

Abbreviations: ABR, auditory brainstem response; 2AFC, two-alternative, forced choice; AIM, auditory image model; BM, basilar membrane; CAP, compound action potential; ERB, equivalent rectangular bandwidth; MPP, masking period pattern; NAP, neural activity pattern; RMS, root mean square; α , exponent determining the rate of change of instantaneous frequency in the chirps; m_+ , m_- , sign for the phase change in ‘Schroeder phase’ harmonic complex tones

times, and it improves the appearance of the compound broadband ABR.

In an effort to increase the magnitude of ABR for clinical research, and to enable frequency-specific stimulation for ABR without using additional high-pass masking noise, Wegner et al. (1997) and Dau et al. (2000) suggested a method for synchronising excitation along the whole length of the cochlea. They created a chirp stimulus with increasing instantaneous frequency that was designed to compensate for the spatial dispersion along the cochlear partition. They showed that these ‘optimised chirps’ enhance ABR wave V, most probably by bringing the low-frequency channels into synchrony with the high-frequency channels. They also showed that playing optimised chirp signals backwards reduces the ABR response markedly, most probably because it increases the phase delays between high- and low-frequency channels up to 20 ms and so desynchronises the low- and high-frequency channels over the full length of the BM. A similar contrast between chirps with increasing instantaneous frequency (up-chirps) and decreasing instantaneous frequency (down-chirps) had previously been reported for the auditory CAP in guinea pigs (Shore and Nuttall, 1985) and for single unit responses from the ventral cochlear nucleus in guinea pigs (Shore et al., 1987).

Despite the phase dispersion, a click stimulus still sounds like ‘click’; that is, a click is, temporally, the most compact sound that we experience. This suggests that auditory processing includes some process that compensates for phase delays, and the physiological data suggest that this process is somewhere between the early neural activity reflected in the ABR, and the representation of sound that we eventually hear, which is more likely in auditory cortex. The sensitivity of the ear to phase differences has been the subject of debate since the early experiments by Ohm in the mid 18th century. He postulated that sound quality is solely dependent on the power spectrum of the sound (Ohm, 1843), and von Helmholtz (1870) performed a set of experiments that seemed to support this conclusion. With the advent of electronic signal generators and headphones, however, it became clear that the ear is not phase deaf (e.g. Mathes and Miller, 1947; see Patterson, 1987 for a review). In a series of experiments with broadband harmonic complex tones (31 equal-amplitude components), Patterson (1987) investigated the extent to which progressive phase delays across a range of channels in the peripheral auditory system are perceptible. When all components are added in cosine phase, the resulting waveform is a click train. Patterson produced chirp-like signals in which the low frequencies appeared first within the cycle by introducing phase changes successively between harmonics and reducing the size of the phase change as harmonic number in-

creased. By varying the rate of change of phase between channels, he attempted to produce a waveform that compensated for the cochlear phase delay. Despite the compensation, these chirps did not produce a sound that was perceived to be more compact than the initial click train. The results were explained using the ‘pulse ribbon’ model of monaural phase perception (Patterson, 1987), which is an early version of the auditory image model (AIM) proposed later (Patterson et al., 1992, 1995). Briefly, it is argued that between-channel phase differences of up to 4 ms (across the frequency region occupied by a signal) are effectively removed at a later stage of auditory processing.

There is no question that the manipulation of the phase spectrum in these experiments has an effect in the cochlea. Smith et al. (1986) and later Kohlrausch and Sander (1995) described masking experiments in which a sinusoidal pip is masked by a complex tone similar to that used by Patterson, but in this case, the component phases were manipulated to produce a flat-envelope waveform (Schroeder, 1970). The instantaneous frequency of the stimulus either increased (m_- Schroeder phase) or decreased (m_+ Schroeder phase) over the course of the cycle. They found that thresholds are considerably reduced over much of the modulation period for the maskers with decreasing instantaneous frequency (m_+ Schroeder phase), relative to both the zero-phase masker and maskers with increasing instantaneous frequency (m_- Schroeder phase). They explained the results using a simulation of BM motion based on the one-dimensional, linear, transmission line model of the cochlea by Strube (1985). The phase condition with decreasing instantaneous frequency (m_+ Schroeder phase) results in deeper valleys between the maxima in BM excitation. They do not, however, describe the perceptions produced by these Schroeder phase waves.

The objective of the current study was to investigate the perception of the ‘optimised’ chirp signals developed by Dau et al. (2000) to manipulate the magnitude of the ABR. Their stimuli differed from those of Patterson (1987) and Smith et al. (1986) insofar as Dau et al. employed a purely time-domain approach, defining the instantaneous frequency of a chirp signal directly from the phase properties of the cochlear partition. All of the chirp signals in the current study had flat envelopes, and since the instantaneous frequency increases exponentially with time, the power spectrum has a low-pass characteristic with a slope of about 6 dB/octave (see figure 1 in Dau et al., 2000). Dau et al. compared ABR results using two forms of chirp stimuli that had the identical phase spectrum: in one case, the waveform had a flat envelope and a low-pass amplitude spectrum; in the other case, the amplitude spectrum was flat and the waveform had an exponential

rising envelope. They found very similar evoked responses to both forms of chirp (figure 8 in Dau et al., 2000) indicating that the precise details of the power spectrum play only a minor role in the synchronisation of neural activity across the cochlea.

Informal listening revealed that up-chirps and down-chirps are readily discriminable, despite their very short duration and their identical power spectra. Three sets of psychoacoustical experiments were performed to assess the detectability of the stimuli, their effectiveness as maskers, and the compactness of their sound quality. The results show that chirp direction does not affect detectability but it does affect masking effectiveness and compactness. The experimental findings are discussed within the framework of a computational model of the peripheral auditory system (Patterson et al., 1992). It is concluded that phase manipulations that produce phase delays between channels have less effect on the perception of chirps than phase manipulations that extend the duration of within-channel fine structure.

2. Optimised chirp signals and BM motion

The synchronising chirp signals described by Dau et al. (2000) are based on the one-dimensional, linear, cochlear model by de Boer (1980). The basic assumption is that BM stiffness $c(x)$ decreases exponentially along the cochlear partition from base to apex, according to:

$$c(x) = C_0 e^{-\alpha x} \quad (1)$$

where $C_0 = 10^4 \text{ N cm}^{-3}$ and the exponent determining the rate of change of instantaneous frequency in the chirps (α) = 3 cm^{-1} for the human cochlea. The mass and damping are assumed to be independent of place, x , and none of the mechanical parameters is level-dependent. The result is an exponential decrease of the speed of the travelling wave along the cochlear partition. With this model and the frequency–place transformation of Greenwood (1990), Dau et al. calculated the instantaneous phase and frequency functions required for a chirp signal that would synchronise excitation along the BM so that all places would reach maximum amplitude at the same moment in time. The resulting function for instantaneous frequency is approximately exponential and the rate of change is determined by the parameter α .

Fig. 1 shows the effect of spatial dispersion along the cochlear partition in a computer simulation of BM motion. In this case, the model is a wave–digital-filter implementation of a one-dimensional transmission line in which stiffness varies with level as well as with place

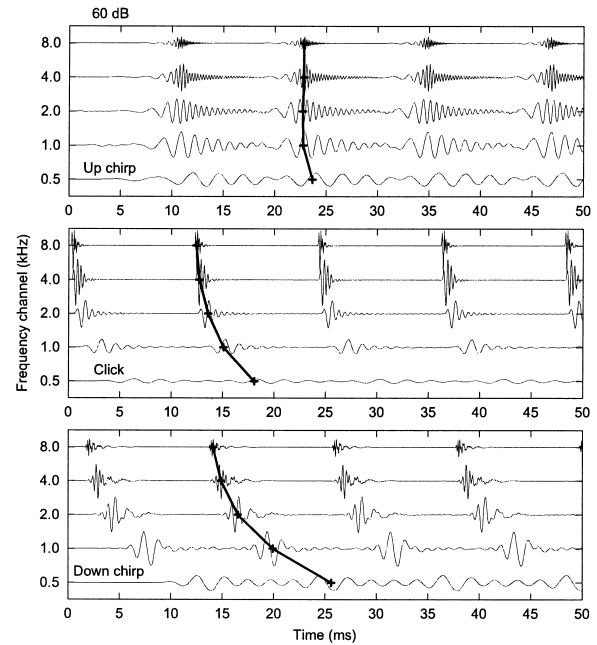


Fig. 1. Computer simulation of BM motion using a one-dimensional, non-linear transmission line. BM velocity in five sections in response to a click (middle panel), an ‘optimised’ chirp with increasing instantaneous frequency (top panel), and the optimised chirp played backwards (bottom panel) is shown. The input level was 60 dB SPL. The solid lines connect the maxima of the Hilbert envelopes in the five channels.

(Giguère and Woodland, 1994). There are four reasons for using this class of model to simulate BM motion rather than a linear, parallel filterbank like the gammatone (Patterson et al., 1995). (1) It enables us to model any effects of level that may arise in the data. (2) The gammatone auditory filter has insufficient temporal asymmetry to explain the differences in the masking produced by m_- and m_+ Schroeder phase waves (see for instance Kohlrausch and Sander, 1995, figures 17 and 18). (3) The impulse response of the gammatone filter does not exhibit the onset chirp observed physiologically (Carney et al., 1999). (4) There is no connection between the channels in a parallel filterbank. In the transmission line filterbank, the coupling of the sections causes a lengthening of the within-channel response to up-chirps which appears to have an important effect on the perception of chirps.

The transmission line filterbank had 100 sections equally spaced on an equivalent rectangular bandwidth (ERB) scale between 70 Hz and 12 kHz. The outputs of five channels are presented in each panel of Fig. 1, with their centre frequencies equally spaced on a logarithmic scale between 500 Hz and 8 kHz. The middle panel shows the responses to a click train; the top panel shows the responses to a sequence of up-chirps; the bottom panel shows the responses to the same chirps played backwards (down-chirps). The solid lines con-

nect the maxima of the Hilbert envelopes in each channel. All of the stimuli were calibrated to a level of 60 dB SPL for the model, and the chirps were repeated at a rate of 83/s (i.e. a 12-ms period).

The responses to clicks in the middle panel show that there is a phase delay between the 500-Hz channel and the 8-kHz channel of approximately 5 ms, reflecting the basic spatial dispersion of the cochlea. The maximum possible delay during click stimulation between base and apex is about 10 ms in the human cochlea. It is this type of figure that prompted the idea of using an up-chirp to align the envelope peaks across channels to produce a ‘super’ click train. The responses to a sequence of up-chirps are shown in the top panel. In this case, it is the ‘optimised’ chirp designed by Dau et al. to compensate for envelope delay. It sweeps from 100 Hz to 10 kHz in 10 ms and the figure shows that it does, indeed, synchronise the moment of maximal excitation for all channels (except for a small deviation in the lowest frequency channel, which is due to a small difference between the models by Giguère and de Boer at low frequencies). Note, however, that in the top panel, the impulse responses within channels are much longer for the up-chirp than for the click. This is a direct result of the coupling between channels of the transmission line filterbank, and it appears that the lengthening is reflected in auditory perception. The bottom panel shows that, when the up-chirp is played backwards, the phase delay between 500 Hz and 8 kHz increases to approximately 12 ms. In this case, however, there is little, or no, lengthening of the impulse response within channel. The shape of the impulse response is a little more complicated than for a click, but the duration of the response is comparable to that of a click.

In summary, the coupling along the cochlear partition imposes a constraint on the internal representation of chirps in the form of a trade-off between the duration of within-channel impulse responses and between-channel phase delays. There is a physical limit to which one can simultaneously decrease the length of the impulse responses and the time delays between sections; this is due to the intrinsic asymmetry of the stiffness function along the cochlea. If we use an up-chirp to align the envelope peaks across channels, it results in a lengthening of the impulse response, a lengthening that does not occur in response to down-chirps. Informal listening revealed that the down-chirp actually sounded more compact than the up-chirp (Uppenkamp and Patterson, 1999). It appears that the length of the impulse response within channel determines the compactness of the perception and the phase delay between channels plays relatively little role in perception despite its dramatic effect on the size of the ABR. The experiments described below were designed to provide a

quantitative assessment of these informal observations.

3. Experiment I: detection threshold for up-chirps and down-chirps

In the first set of experiments, detection threshold was measured in quiet and in the presence of a broadband noise to determine whether the different excitation of up-chirps and down-chirps produces a difference in threshold, and to determine whether threshold was simply determined by the total energy of the stimulus.

3.1. Experimental procedures

Detection threshold was measured using a two-interval, two-alternative, forced choice (2AFC) adaptive procedure. Both intervals had a duration of 720 ms. The intervals were marked by warning lights and separated by a 500-ms inter-stimulus interval. The stimulus repetition period was an experimental parameter, varying between 30 and 240 ms, and the number of single chirps in each stimulus interval varied accordingly. Starting at about 30 dB above threshold, the level was adaptively varied using a 3-down/1-up rule until 12 reversals were completed. Threshold was defined to be the mean of the levels at the last eight reversals. The step size was 5 dB during the first four reversals and it was reduced to 2 dB thereafter.

For the masked-threshold conditions, the masker was a uniformly exciting noise. The slope of the power spectral-density function was calculated according to the formula given by Glasberg and Moore (1990) for roex auditory filters equally spaced on an ERB scale. The masker was presented at three levels: 58, 38 and 18 dB SL. These sensation levels were determined in advance by measurement of absolute threshold for the noise masker with two normal-hearing subjects. The test signal was either one single up-chirp/down-chirp or sequences of up-chirps/down-chirps presented at repetition periods of 100 and 20 ms, respectively. All test signals were temporally centred in the masker.

3.2. Equipment and listeners

All stimuli were generated digitally at a sampling frequency of 25 kHz on a personal computer. They were played via a TDT system II, using 16-bit D/A converters, an anti-aliasing filter with a cut-off frequency of 10 kHz, programmable attenuators, a weighted summer, and a headphone buffer. The stimuli were presented diotically via headphones AKG 240-D. The listeners were seated in a double-walled, sound-insulating booth (IAC). They responded via a four-button

box controlled by the parallel interface of the TDT system. Feedback was provided with lights at the end of each trial, during all of the detection experiments. Thresholds in quiet and masked thresholds were determined for three normal-hearing listeners (aged between 23 and 36 years) in each experiment; two of them participated in both experiments.

3.3. Results

3.3.1. Threshold in quiet

Fig. 2 shows detection threshold for up-chirps and down-chirps as a function of stimulus repetition period from 30 to 240 ms. The parameter is sweep rate which determines chirp duration and total energy; the data are for the sweep rate parameters $\alpha=2.0$ (3.1 ms duration, open symbols) and $\alpha=3.0$ (10.5 ms duration, filled symbols). Threshold estimates were obtained for each experimental condition; the data points in the figure show the overall means for three listeners with standard errors over listeners. For both sweep rates, threshold for up-chirps and down-chirps is essentially the same; the differences are typically less than 1 dB and none of the differences is statistically significant. This indicates that differences in supra-threshold sound quality are not relevant for detection of the chirps in silence. The bold solid line in Fig. 2 shows the slope of the function relating threshold to repetition period, based on linear regression to all of the data points. It shows that threshold increases 1.8 dB per doubling of the repetition period, i.e. per halving the number of chirps in the sequence.

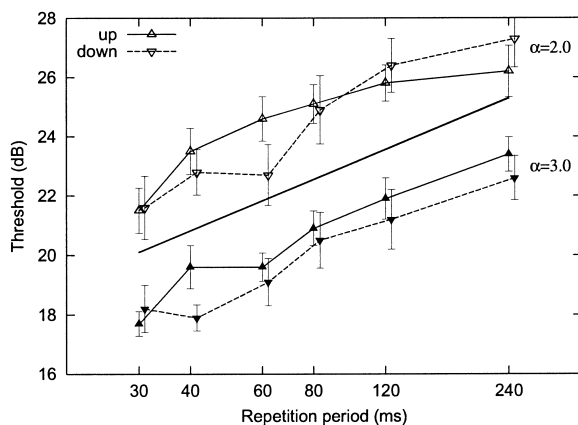


Fig. 2. Detection threshold in quiet for sequences of chirps as a function of repetition period. Mean values with standard errors are shown for three listeners for up-chirps and down-chirps at two different sweep rates (α values of 2.0 and 3.0). The bold solid line shows the linear regression of threshold on repetition period for all data points. The slope is 1.8 dB per doubling of the repetition period.

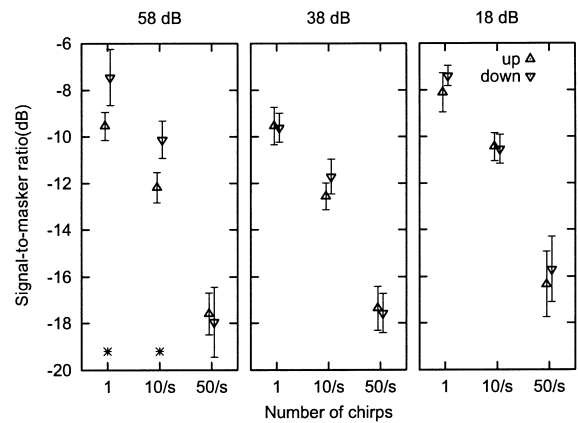


Fig. 3. Detection threshold in uniformly exciting noise for up-chirps and down-chirps. Mean results for three listeners at three different noise levels. The asterisks mark statistically significant differences between up- and down-chirps; the bars indicate standard errors.

3.3.2. Masked thresholds in uniform exciting noise

Detection thresholds for chirps masked by uniformly exciting noise were measured for single chirps and sequences of chirps at repetition rates of 10 s^{-1} and 50 s^{-1} . The masker level was 58, 38 or 18 dB SL. Three subjects did three runs for up-chirps and down-chirps in all conditions. Fig. 3 shows the results averaged across listeners with the standard errors. There was no overall difference between the detectability of up-chirps and down-chirps. Specific comparisons showed a small but significant difference (2 dB) at the highest masker level (58 dB SL) for single chirps and chirps repeated at 10 s^{-1} . This may indicate that the up-chirp is marginally more detectable when the chirps are perceived as single events.

3.4. Discussion

The fact that detection threshold is essentially the same for up-chirps and down-chirps presented in silence and in noise suggests that threshold might be determined by the overall energy of the chirp. However, this conclusion does not generalise to the other variables. Fig. 2 shows that there is an increase of 1.8 dB per doubling of stimulus repetition period; each doubling effectively halves the overall stimulus energy and so complete energy integration would result in a 3-dB increase in threshold per halving of the number of chirps. The slope of 1.8 dB is just a little over the 1.5 dB per halving that would be predicted by statistical combination of independent events. For example, consider the ‘multiple looks’ model of Viemeister and Wakefield (1991) which assumes that the output of the auditory filter is integrated with a short time constant ($\sim 3 \text{ ms}$) which would preserve each chirp as a separate event in the current experiment. It is argued that threshold rises with repetition period, not because the total energy de-

creases, but because there are fewer opportunities to sample the process ('looks') in a statistical sense. Decisions can be based on selected looks. Such models predict an increase in threshold of 1.5 dB halving of the number of separate items in the sequence. However, this cannot be the complete explanation either: The average distance between the data points for the two sweep rates ($\alpha=2.0$ and $\alpha=3.0$) is about 4 dB, whereas the overall change in root mean square (RMS) level is 5.7 dB, and the number of chirps does not vary with sweep rate. A reasonable hypothesis appears to be that there is integration of energy for individual chirps but that it is incomplete for the longer chirps, and that multiple looks accounts for the increase in threshold with decreasing numbers of events.

Since the audiogram is not flat and our stimuli have relatively more energy at lower frequencies, the signal is probably most detectable in the octave from 500 to 1000 Hz. It may be that this restriction in the frequency range also contributes to the lack of differences in the detectability of up- and down-chirps.

The fact that there is a statistically significant advantage for up-chirps at the highest masker level suggests that there may be a small difference in perceived loudness for up- and down-chirps, once they are well above absolute threshold. However, loudness-matching experiments with very short stimuli like these are notoriously difficult, so it would be difficult to confirm. In any event, it seems unlikely that this small difference would play an important role in the remaining experiments.

4. Experiment II: masking period patterns (MPPs)

In the second experiment, MPP were measured using tone pips as test signals and a sequence of chirps as the masker. These MPP provide a means of investigating differences in the fine structure of the BM motion evoked by up-chirps and down-chirps. It is assumed that the pip threshold reflects the temporal course of the frequency sweep evoked by the masker (Zwicker, 1976; Kohlrausch and Sander, 1995).

4.1. Procedure

The maskers were sequences of up-chirps and down-chirps with $\alpha=3.0$ (10.5-ms duration); the repetition period was 20 ms. Each chirp in the sequence was gated individually using a Hanning window with a 1-ms rise-time and a 1-ms fall-time to reduce the impact of sharp onsets and offsets. The stimulus intervals had a duration of 500 ms and they were separated by a 250-ms inter-stimulus interval. The signal was a tone pip composed of two cycles of either a 1-kHz tone (2 ms duration) or a 4-kHz tone (0.5-ms duration), and they were

gated with a Hanning window with the same duration. The masker level was 45 dB above threshold and it was constant throughout the run. The initial difference between masker and signal attenuation was 10 dB and a two-interval 2AFC adaptive procedure was used. The signal level was changed adaptively using a 3-down/1-up algorithm. The delay between the signal and masker ranged from +10 ms to -10 ms relative to the middle of the masker in 2 ms steps. The order of these delays was randomised and three threshold estimates were collected for each condition and every listener.

The equipment was the same as in Experiment I. There were three normal-hearing listeners, two of whom (SF and SU) had participated in Experiment I.

4.2. Results and discussion

Fig. 4 shows the average thresholds for the three listeners (with standard deviations) as a function of the delay of the signal relative to the middle of the masker. The left column is for the 1-kHz tone pip; the right column is for the 4-kHz tone pip. The top and middle rows show the MPPs for up-chirps and down-chirps, respectively. The dotted lines show the temporal position of the masker relative to the delay axis. Since the chirps were repeated with a period of 20 ms, the conditions where the delay is +10 ms and -10 ms are the same. The bottom row shows the up and down patterns plotted together, with the down-chirp data reversed in time to enable direct comparison of thresholds associated with a specific instantaneous frequency in the masker. The threshold levels are presented as signal-to-masker ratios in dB. Conditions where the difference between the thresholds in the up-chirp and down-chirp maskers were statistical significant are indicated by asterisks.

Thresholds obtained with the 1-kHz signal ranged from -12 dB to +10 dB with the maxima occurring at the 4-ms delay. Significant differences between the up-chirp and down-chirp maskers occur towards the high-frequency end of the chirp. It is always the up-chirp masker which produces more masking and the difference can be as large as 8 dB. These differences probably reflect the difference in the BM response to up-chirps and down-chirps (cf. Fig. 1). In each channel, the excitation produced by down-chirps is more compact in time than that produced by up-chirps. This leaves deeper valleys between the peaks of excitation, especially for frequencies below 2 kHz, and this makes the sequence of down-chirps a less effective masker. The listeners were apparently able to detect the part of the signal excitation in the valleys.

These results are similar to the findings reported by Kohlrausch and Sander (1995) for MPPs obtained with harmonic complex maskers. They used m_- and m_+

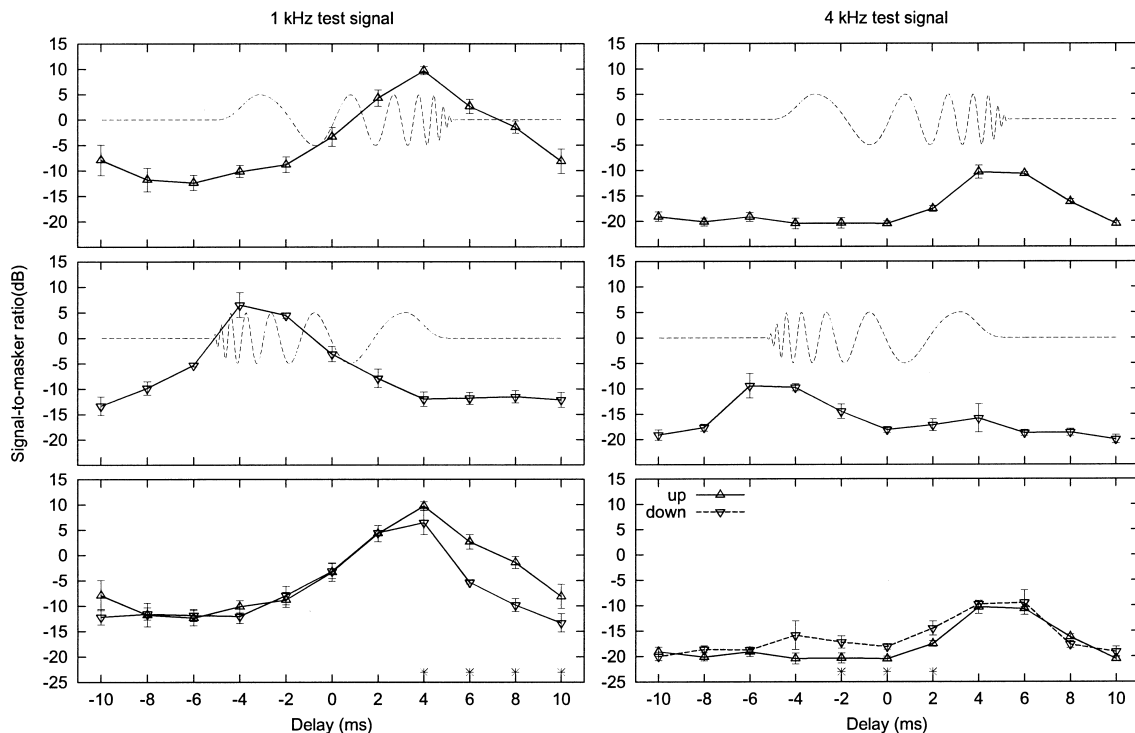


Fig. 4. MPPs for tone pips (solid lines) masked by sequences of up-chirps and down-chirps at a repetition period of 20 ms. Threshold is presented as a function of the position of the tone pip relative to the masker (dashed lines). Left column: 1-kHz signal (two cycles, 2-ms duration). Right column: 4-kHz signal (two cycles, 0.5-ms duration). Mean results for three listeners. Top row: up-chirp masker (Δ); middle row: down-chirp masker (∇); bottom row: data from the top and middle panels, with the down-chirp results reversed in delay in order to compare the thresholds for a given instantaneous frequency. The asterisks mark statistically significant differences in threshold; the bars show the standard errors.

Schroeder phase waves, which are, effectively, forms of up-chirps and down-chirps. They found that the two phase conditions could produce differences in threshold of up to 25 dB in the minima of the MPPs. They also found that the down-chirp (m_+ Schroeder phase) produced less masking. However, both their maskers and signals were physically different from ours and it is difficult to make a quantitative comparison of the data.

Thresholds obtained with the 4-kHz signal ranged from -21 dB to -10 dB; that is, the chirp sequence is a much less effective masker for the higher frequency signal, and the range of the MPP is considerably smaller than for the 1-kHz signal. The overall reduction in threshold at 4 kHz reflects the low-pass characteristic of the power spectrum of the chirps caused by the exponential increase of instantaneous frequency with time in this study. There are significant differences between pairs of up- and down-chirps for delays between -2 and 2 ms. However, they are much smaller than the differences with the 1 kHz signal (3 dB or less), and they are smaller than the non-significant difference at -4 ms. It is also the case that the differences occur at positions close to the middle of the chirp, corresponding to instantaneous frequencies below 1 kHz, and

threshold is higher for the down-chirps rather than the up-chirps. If these differences represent a replicable effect, it is not clear within the current framework, what that effect is.

5. Experiment III: the perceptual compactness of up-chirps and down-chirps

The third experiment was designed to quantify the asymmetry in perceived sound quality of up-chirps and down-chirps. Clicks and chirps with different sweep rates were compared in terms of the ‘compactness’ of the perception using the method of paired comparisons.

5.1. Procedure

Two groups of five listeners participated in this experiment. For one group, the mean presentation level was held at 40 dB HL during the experiment and there were 24 chirps (12 different sweep rates for the two directions, up and down). The sweep rate parameter, α , was varied from 3.5 to 0.1, resulting in a chirp duration from 20.2 ms to 0.4 ms in 1.8-ms steps. A click stimulus was included in the set, a rectangular pulse of

80- μ s duration (two samples). All stimuli were scaled to have the same RMS value. Each stimulus interval contained a sequence of three identical stimuli repeated at a rate of 4.2 Hz, which is well below the pitch range. During a trial, two different stimulus sequences were presented with an inter-stimulus interval of 500 ms. A roving level paradigm was used to minimise the influence of possible differences in loudness. The presentation level was randomly varied over a range of ± 3 dB between intervals within each trial. The listener's task was to choose the interval that sounded more compact or more like a click. Over the course of the experiment, the listeners heard all possible combinations of the stimuli, and in both orders. Before testing began, the listeners were presented a series of clicks and told that they were clicks and that this sound was what we meant by a compact sound. Each trial could be repeated once by the listener before they were forced to give a response.

The second group of listeners did a version of this paired comparison experiment with fewer stimuli but presented at three different levels in an interleaved design, to test for the influence of presentation level. The task was the same and the basic design was very similar. Thirteen of the previous set of 25 stimuli were used: six up-chirps and six down-chirps at durations of 2.2, 5.8, 9.4, 13.0, 16.6 and 20.2 ms, and the click stimulus. This group did all possible comparisons at three different presentation levels of 20 dB HL, 40 dB HL, and 60 dB HL. Comparisons across different levels did not occur within a trial.

Prior to the main experiment, a reduced version involving nine stimuli from across the full range was run to familiarise the listeners with the response box and the procedure. This training set had 72 trials. During

the full experiment, every possible combination was presented twice, once as A–B and once as B–A. No stimulus was compared to itself, so the complete experiment had $25 \times 24 = 600$ trials per listener for group 1 and $3 \times 13 \times 12 = 478$ trials for group 2. The order of trials was randomised, which interleaved the three presentation levels in the case of group 2. To avoid the reduction of attention associated with long runs, the trials were then subdivided into 24 runs with 25 trials each for group 1, and 18 runs with 26 trials each, for group 2. The listeners were advised to take a break every four runs to ensure they maintained concentration.

The equipment was the same as that used in Experiments I and II, except that a low-pass filter with a cut-off frequency of 8 kHz was added to minimise the influence of onset and offset transients. No feedback was given. Two of the ten listeners in this experiment participated in the detection experiments. The remaining listeners had no previous experience with psychoacoustical tasks. They were aged between 16 and 36 years.

5.2. Results and discussion

A relative scale of 'compactness', reflecting the contrasts between the stimuli, was constructed from the data of the paired comparison experiment using the Bradley–Terry–Luce method (David, 1988). The primary assumption is that the stimuli can be ordered according to a linear scale. The judgements from all five listeners in group 1 were pooled to produce 10 observations for each pair of stimuli (five with order A–B and five with order B–A) – a total of 3000 trials, or 240

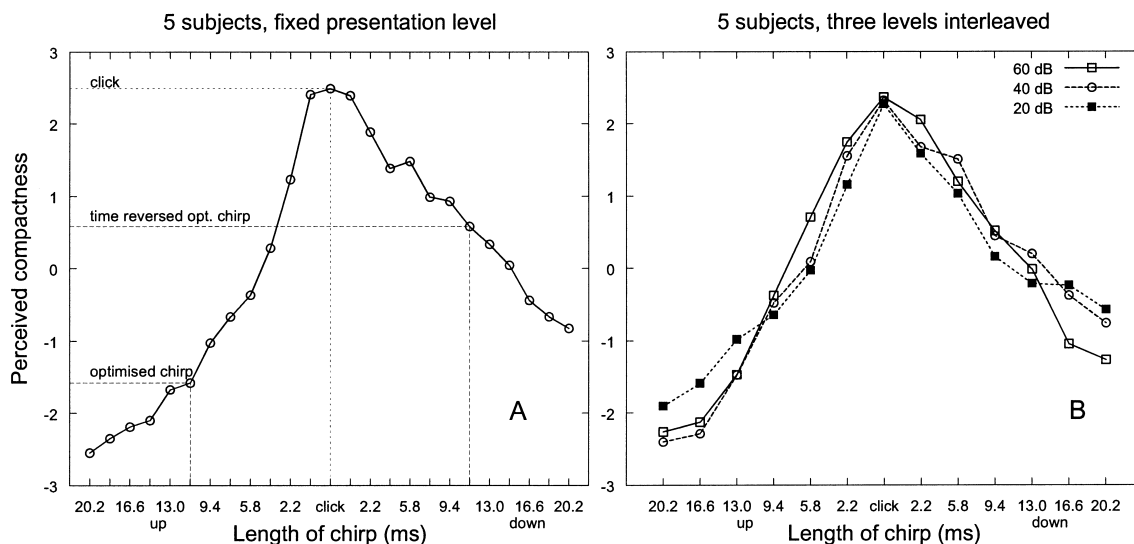


Fig. 5. 'Compactness' as a function of sweep rate (chirp length) derived from a paired comparison experiment with up- and down-chirps. The judgements of the five listeners were pooled for each function. A: one stimulus level (40 dB SL) and many sweep rates. B: Fewer sweep rates and three interleaved levels (20, 40 and 60 dB).

observations for each stimulus. Fig. 5A shows the resulting scale of ‘compactness’ as a function of the duration of the chirp signals; the up-chirps are on the left, the click is in the middle and the down-chirps are on the right. The ordinate covers a range of approximately five points (–2.5 to 2.5). It is a relative scale, that is, only differences are relevant. The zero-line is intrinsic to the analysis and has no particular meaning.

The obvious and most important finding is the pronounced asymmetry between up-chirps and down-chirps, and the fact that it is the down-chirp that sounds more like a click, for all stimulus durations longer than 2 ms. The contrast between the compactness of the optimised chirp, that is, the chirp designed to compensate for spatial dispersion along the BM, and the time-reversed optimised chirp is a full 2.1 points on this 5-point scale (the vertical displacement between the dotted lines). This is despite the synchronisation of all frequency channels by the optimised chirp, and the enhancement of the phase delay by up to 20 ms across channels by the time-reversed optimised chirp. It confirms the hypothesis that the perception of short frequency sweeps is more affected by within-channel fine structure than by between-channel phase differences. This contrasts with the increase of the amplitude of ABR wave V during stimulation with optimised chirp signals reported by Dau et al. (2000).

Fig. 5B shows the results for group 2 where three different levels were interleaved. As previously, the data of all five listeners were analysed together. The individual functions are somewhat more variable because of the reduced number of observations ($5 \times 13 \times 12 = 780$ trials per function, rather than a total of 3000 in Fig. 5A). Nevertheless, the asymmetry in the perceived compactness of up-chirps and down-chirps remains and it exists even when the level is only 20 dB HL. Moreover the shape of the function does not vary systematically with level. The three functions in Fig. 5B should be read as separate relative scales that cannot be directly compared, since there were no cross-level judgments within trials, and the data in the analysis that produce each function are all from one specific level. So, absolute differences in perceived compactness between levels, and the range for each function, is not actually represented by these functions. Such a comparison would require a much larger experiment that included all possible cross-level comparisons ($39 \times 38 = 1482$ trials per listener), which cannot be completed in a single session.

The data from both experiments also show that none of the chirp stimuli was perceived to be more compact than the click stimulus. There is essentially no difference in perceived compactness between the click and the shortest duration chirps (0.4 ms) for group 1. The 0.4-ms stimuli sound like they would be just discrimi-

nable from the click but neither is perceived to be more compact. The finding that the true click is perceived as the most compact stimulus in the set is not surprising since the click was used to illustrate compactness to the listeners at the beginning of the experiment. Nor is it crucial to the primary issue in this study, which is the contrast between the effect that chirp direction has on the ABR and on perception. Nevertheless, informal listening and the consistency of the data across listeners suggests that the click does have a unique position on the compactness scale, confirming the original finding that channel alignment does not produce a ‘super click’ (Patterson, 1987).

The finding that the perceived compactness of chirps is not increased by the stimulus that produces alignment of envelope peaks in the cochlea and an increase in the ABR (Dau et al., 2000) indicates that perceived compactness of these chirps is apparently not determined by channel alignment. It also indicates that an additional stage of processing involving temporal integration intervenes between the representation at the input to the inferior colliculus and the representation upon which our perceptions are based. One potential model of such auditory processing is presented in the Section 6.

6. Auditory representations of clicks and chirps

The AIM (Patterson et al., 1992) is a time-domain simulation of the early stages of processing in the auditory system, including BM motion, neural transduction, and temporal integration. AIM was used to simulate the response to up-chirps and down-chirps to investigate the effect of asymmetry in the model and determine whether it could explain both the ABR results and the perceptual results. The outputs of the first two stages of AIM are clearly different for the optimised chirp and its time-reversed counterpart, and in a way that might explain the change in the ABR. The question is (a) whether the final stage, which produces the auditory image, reduces or eliminates the differences when there is no perceived difference in compactness, and (b) whether some feature in the auditory image of the optimised chirp is in some way less compact than the corresponding feature in the time-reversed version of the optimised chirp.

The simulations were performed with the ‘physiological route’ in AIM which is described in Patterson et al. (1995). It uses the non-linear, transmission line model of BM motion of Giguère and Woodland (1994), as before. Neural transduction is simulated with one inner haircell per channel, using the model of Meddis (1988) for a medium spontaneous-rate fibre. Finally, ‘strobed’ temporal integration is applied to the neural activity

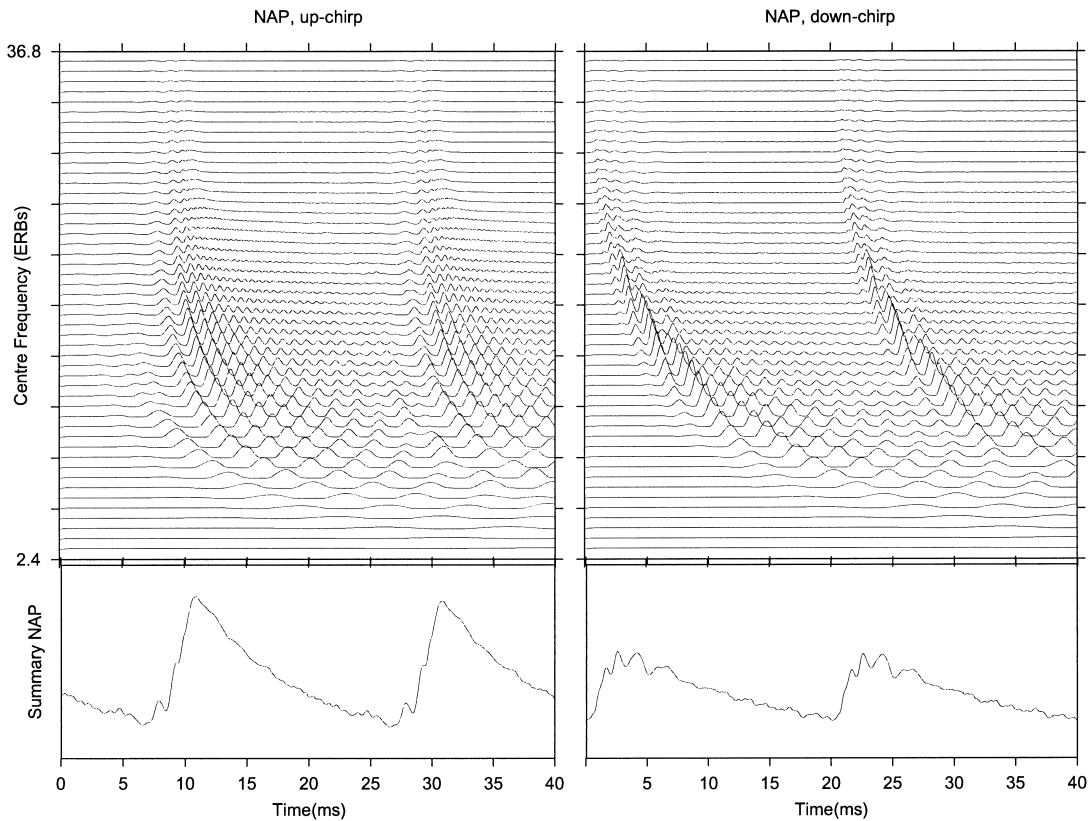


Fig. 6. Computer simulation of the NAPs for up-chirps (left) and down-chirps (right), repeated at a rate of 50 Hz. A non-linear, one-dimensional, transmission line model with 100 sections was used to produce the BM motion, and the inner haircell model of Meddis (1988) was used to transform the BM motion into a NAP. The stimulus level was 55 dB SPL. The bottom panels show summary NAPs created by averaging the patterns across all frequency channels.

pattern (NAP) to transform it into a stabilised auditory image. This temporal integration mechanism effectively removes between-channel phase differences while preserving within-channel fine structure.

The BM velocity produced in response to clicks, up-chirps and down-chirps was presented in Fig. 1 for five sections of the transmission line. The up-chirp aligns the point of maximal excitation across channels, but it also increases the length of the response within channels. The result is that all sections of the transmission line are driven by the acoustic waveform simultaneously up to their respective characteristic frequencies, and then each channel rings at its characteristic frequency for a certain duration. Down-chirps are much less effective in evoking within-channel ringing, but there is a large delay between the point of maximum excitation across channels. This basically reflects the intrinsic temporal asymmetry of the place–frequency mapping in the cochlea where the high-frequency channels are excited first.

The top panels in Fig. 6 show the transformation of the BM motion produced by the up-chirps and down-chirps into NAP. Fifty sections of the transmission line are presented to illustrate the full surface of activity.

The stimulus repetition period is 20 ms for this, and subsequent, simulations. The haircell model reduces the contrast between up and down because it includes half-wave rectification, low-pass filtering and adaptation. Nevertheless, the main differences are preserved at this stage; the up-chirp produces synchronisation across channels and longer responses within channels. The synchronisation results in larger peaks when the neural activity is averaged across frequency channels, as is illustrated in the bottom panels of Fig. 6. It is this effect which produces the increase of ABR wave V in response to up-chirps.

The next stage in AIM is strobed temporal integration, which transforms the flowing NAP into a stabilised representation. Broadly speaking, for each channel of the NAP, time-intervals between the largest peaks in the NAP and all other peaks in the fine structure of the NAP are extracted and stored in a time-interval histogram; the details are presented in Patterson (1994a,b). The auditory image decays exponentially at each point in the image with a half-life of 30 ms and this relatively long time constant makes this the dominant form of temporal integration in AIM. The top panels in Fig. 7 illustrate the effect of strobed temporal integration for a

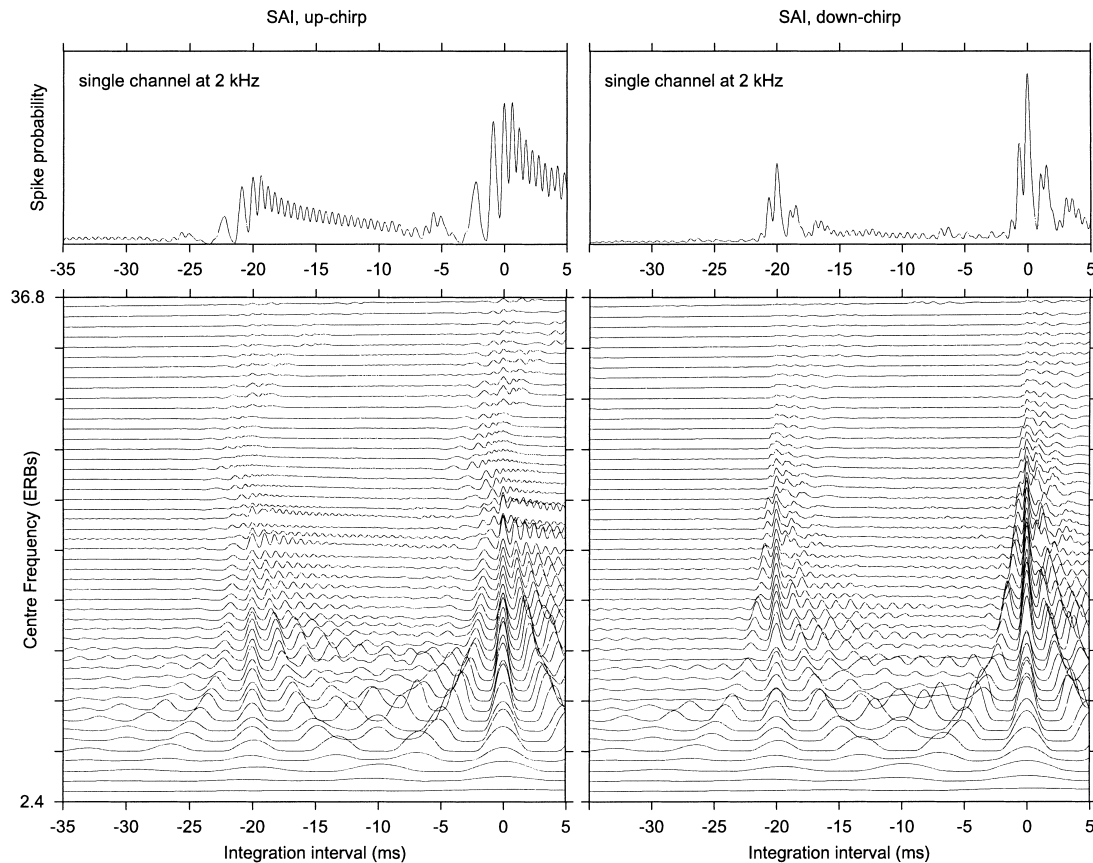


Fig. 7. Stabilised auditory images for up-chirps (left) and down-chirps (right). The top panels illustrate the effect of strobed temporal integration in the frequency channel centred at 2 kHz (21.1 ERBs). The bottom panels show the complete auditory images, using all of the channels of the NAPs in Fig. 6.

single frequency channel centred on 2 kHz. Temporal integration has taken place at this point, but the introduction of the time-interval dimension means that the fine structure of this periodic sound is preserved. Comparison of the panels shows how the up-chirp has extended the impulse response relative to the down-chirp. The complete array of time-interval histograms is referred to as the auditory image and it is stabilised whenever the sound is periodic because the auditory image is continuously updated from the input NAP. The bottom panels in Fig. 7 show the auditory images derived from extended versions of the NAPs in Fig. 6. The periodicity of the stimulus is revealed by the ridge of peaks across channels at the -20 -ms integration interval. These images for up-chirps and down-chirps are more similar than their NAPs because the conversion to the time-interval representation introduces a form of phase alignment. A careful inspection, however, of the main ridges at 0 ms and -20 ms integration interval and the adjacent ridges, shows that the activation is more compact in the case of the down-chirps. The peaks are larger along the main ridge in all channels and there is less activity to the right of the ridge at -20 ms. The differences in within-channel fine structure, which were

apparent in the BM motion, are preserved in this stabilised representation.

7. Summary and conclusions

The demonstration by Dau et al. (2000) that the direction of the frequency sweep in a brief chirp can markedly increase or decrease the magnitude of the ABR led us to investigate the psychophysical differences between up-chirps and down-chirps; specifically, it led us to compare their detectability, their effectiveness as maskers, and the compactness of their sound quality. The main findings can be summarised as follows: (1) The direction of the frequency sweep does not affect the detection of up- and down-chirps, which have the same rate of sweep and the same energy. This does not, however, mean that the energy of these chirps is simply integrated over time and frequency. Firstly, as the rate of sweep decreases, threshold decreases because the duration and the energy of these fixed-bandwidth chirps increase. The decrease, however, is only about $2/3$ of that which would be expected from the energy increase. Secondly, as the number of chirps in a sequence in-

creases, thresholds decrease, but by about 1.5 dB per doubling of the number of chirps rather than 3 dB per doubling. This is more in line with the ‘multiple looks’ model of Viemeister and Wakefield (1991) rather than energy integration. (2) The differences in the BM motion produced by up-chirps and down-chirps do affect the MPPs that they produce. The concentrated motion produced in response to down-chirps at the start of the period is followed by deeper valleys later in the period and thus threshold is lower over a portion of the cycle. (3) There is a big difference in the perceived sound quality of up-chirps and down-chirps. Down-chirps are perceived as more compact, or more click-like, than up-chirps. This occurs despite the synchronisation of activity across frequency channels produced by up-chirps, and the extended phase delay produced by down-chirps. The sound quality of short frequency sweeps appears to reflect within-channel fine structure rather than between-channel phase differences.

Time-domain models of auditory perception (Meddis and Hewitt, 1991; Patterson et al., 1992) which convert the phase-locked NAP flowing from the cochlea into an array of time-interval histograms can explain these effects, at least, qualitatively, because the time-interval calculation removes between-channel phase differences. The apparent contrast between the effect of sweep direction on the ABR and the perception of compactness suggests that the temporal integration that removes between-channel phase differences is located beyond the input to the inferior colliculus where wave V of the ABR is thought to be generated.

Acknowledgements

This work was supported by the UK Medical Research Council (G9703469, G9901257). Thanks to Torsten Dau, Andre Rupp and Birger Kollmeier for many fruitful discussions and helpful suggestions, and to Daniel Pressnitzer for advice and a routine to perform the Bradley–Terry–Luce analysis of the paired comparison data.

References

de Boer, E., 1980. Auditory physics. physical principles in hearing theory I. *Phys. Rep.* 62, 87–174.

Carney, L.H., McDuffy, M.J., Shekhter, I., 1999. Frequency glides in the impulse responses of auditory-nerve fibers. *J. Acoust. Soc. Am.* 105, 2384–2391.

Dau, T., Wegner, O., Mellert, V., Kollmeier, B., 2000. Auditory brainstem responses (ABR) with optimized chirp signals compensating basilar membrane dispersion. *J. Acoust. Soc. Am.* 107, 1530–1540.

David, H.A., 1988. The method of paired comparisons. 2nd edn., Oxford University Press, New York.

Don, M., Eggermont, J.J., 1978. Analysis of click-evoked brainstem potentials in man using high-pass noise masking. *J. Acoust. Soc. Am.* 63, 1084–1092.

Don, M., Ponton, C.W., Eggermont, J.J., Masuda, A., 1994. Auditory brainstem response (ABR) peak amplitude variability reflects individual differences in cochlear response times man using high-pass noise masking. *J. Acoust. Soc. Am.* 96, 3476–3491.

Glasberg, B.R., Moore, B.C.J., 1990. Derivation of auditory filter shapes from notched noise data. *Hear. Res.* 47, 103–108.

Giguère, C., Woodland, P.C., 1994. A computational model of the auditory periphery for speech and hearing research I, ascending path. *J. Acoust. Soc. Am.* 95, 331–342.

Greenwood, D.D., 1990. A cochlear frequency-position function for several species – 29 years later. *J. Acoust. Soc. Am.* 87, 2592–2605.

von Helmholtz, H., 1870. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik.* 3. Ausgabe. Vieweg, Braunschweig.

Johnstone, J.R., 1981. Origin of the guinea pig cochlear action potential produced by a click. *Hear. Res.* 4, 347–352.

Kohlrausch, A., Sander, A., 1995. Phase effects in masking related to dispersion in the inner ear. II. Masking period patterns of short targets. *J. Acoust. Soc. Am.* 97, 1817–1829.

Mathes, R.C., Miller, R.L., 1947. Phase effects in monaural perception. *J. Acoust. Soc. Am.* 19, 780–797.

Meddis, R., 1988. Simulation of auditory-neural transduction: further studies. *J. Acoust. Soc. Am.* 83, 1056–1063.

Meddis, R., Hewitt, M.J., 1991. Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity. *J. Acoust. Soc. Am.* 89, 2883–2894.

Ohm, G.S., 1843. Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Einrichtungen. *Ann. Phys. Chem.* 59, 512–565.

Patterson, R.D., 1987. A pulse ribbon model of monaural phase perception. *J. Acoust. Soc. Am.* 82, 1560–1586.

Patterson, R.D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., Allerhand, M., 1992. Complex sounds and auditory images. In: Cazals, Y., Demany, L., Horner, K. (Eds.), *Auditory Physiology and Perception.* Pergamon, Oxford, pp. 429–446.

Patterson, R.D., 1994a. The sound of a sinusoid: spectral models. *J. Acoust. Soc. Am.* 96, 1409–1418.

Patterson, R.D., 1994b. The sound of a sinusoid: time-interval models. *J. Acoust. Soc. Am.* 96, 1419–1428.

Patterson, R.D., Allerhand, M.H., Giguère, C., 1995. Time-domain modelling of peripheral auditory processing: a modular architecture and a software platform. *J. Acoust. Soc. Am.* 98, 1890–1894.

Schroeder, M.R., 1970. Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans. Inf. Theory* 16, 85–89.

Shore, S.E., Nuttall, A.L., 1985. High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts. *J. Acoust. Soc. Am.* 78, 1286–1295.

Shore, S.E., Clopton, B.M., Yolande, N., 1987. Unit responses in ventral cochlear nucleus reflect cochlear coding of rapid frequency sweeps. *J. Acoust. Soc. Am.* 82, 471–478.

Smith, B.K., Sieben, U.K., Kohlrausch, A., Schroeder, M.R., 1986. Phase effects in masking related to dispersion in the inner ear. *J. Acoust. Soc. Am.* 80, 1631–1637.

Strube, H.W., 1985. A computationally efficient basilar-membrane model. *Acustica* 58, 207–214.

Uppenkamp, S., Patterson, R.D., 1999. Auditory representations of optimised chirp signals. In: Dau, T., Hohmann, V., Kollmeier, B. (Eds.), *Psychophysics, Physiology and Models of Hearing.* World Scientific, Singapore, pp. 235–239.

- Uppenkamp, S., Patterson, R.D., Rupp, A., Scherg, M., Dau, T., 2001. The neural basis of the perception of temporal asymmetry in short frequency sweeps. In: Breebart, D.J., Houtsma, A.J.M., Kohlrausch, A., Prijs, V.F., Schoonhoven, R. (Eds.), *Physiological and Psychophysical Bases of Auditory Function – 12th International Symposium on Hearing*. Shaker, Maastricht, pp. 343–351.
- Viemeister, N.F., Wakefield, G.H., 1991. Temporal integration and multiple looks. *J. Acoust. Soc. Am.* 90, 858–865.
- Wegner, O., Dau, T., Uppenkamp, S., 1997. Auditory brainstem responses with optimized chirp signals compensating BM dispersion. In: Schick, A., Klatt, M. (Eds.), *7th Oldenburg Symposium on Psychoacoustics*. BIS, Oldenburg, pp. 83–90.
- Zwicker, E., 1976. Masking period patterns of harmonic complex tones. *J. Acoust. Soc. Am.* 60, 429–439.