Reaction times reflect subjective auditory perception of tone sequences in macaque monkeys

Article *in* Hearing research · September 2012

DOI: 10.1016/j.heares.2012.08.014 · Source: PubMed

6 authors, including:

Elena Selezneva
Leibniz Institute for Neurobiology
16 PUBLICATIONS **487** CITATIONS
SEE PROFILE

Alexander Georgievich Gorkin
Russian Academy of Sciences
30 PUBLICATIONS **93** CITATIONS
SEE PROFILE

Judith Mylius
German Primate Center
12 PUBLICATIONS **48** CITATIONS
SEE PROFILE

Toemme Noesselt
Otto-von-Guericke-Universität Magdeburg
60 PUBLICATIONS **2,413** CITATIONS
SEE PROFILE

Hearing Research 294 (2012) 133–142

Contents lists available at SciVerse ScienceDirect

Hearing Research

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

Research paper

Reaction times reflect subjective auditory perception of tone sequences in macaque monkeys

Elena Selezneva [a,*], Alexander Gorkin [a,b], Judith Mylius [a], Tömme Noesselt [c], Henning Scheich [a], Michael Brosch [a]

[a] Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany
[b] Institute of Psychology, Russian Academy of Sciences, 129366 Moscow, Russia
[c] Institute of Psychology II, Otto-von-Guericke University, 39120 Magdeburg, Germany

ARTICLE INFO

Article history:
Received 24 April 2012
Received in revised form
10 August 2012
Accepted 29 August 2012
Available online 16 September 2012

ABSTRACT

Perceptually ambiguous stimuli are useful for testing psychological and neuronal models of perceptual organization, e.g. for studying brain processes that underlie sequential segregation and integration. This is because the same stimuli may give rise to different subjective experiences. For humans, a tone sequence that alternates between a low-frequency and a high-frequency tone is perceptually bistable, and can be perceived as one or two streams. In the current study we present a new method based on response times (RTs) which allows identification ambiguous and unambiguous stimuli for subjects who cannot verbally report their subjective experience. We required two macaque monkeys (*macaca fascicularis*) to detect the termination of a sequence of light flashes which were either presented alone, or synchronized in different ways with a sequence of alternating low and high tones. We found that the monkeys responded faster to the termination of the flash sequence when the tone sequence terminated shortly before the flash sequence and thus predicted the termination of the flash sequence. This RT gain depended on the frequency separation of the tones. RT gains were largest when the frequency separation was small and the tones were presumably heard mainly as one stream. RT gains were smallest when the frequency separation was large and the tones were presumably mainly heard as two streams. RT gain was of intermediate size for intermediate frequency separations. Similar results were obtained from human subjects. We conclude that the observed RT gains reflect the perceptual organization of the tone sequence, and that tone sequences with an intermediate frequency separation, as for humans, are perceptually ambiguous for monkeys.

1. Introduction

Perceptually ambiguous stimuli are helpful tools in understanding how the brain processes sensory stimuli and for testing psychological models of perceptual organization. This is because the same stimulus gives rise to different percepts at different times. Thus, neuronal processes that correlate with stimulus features can be distinguished from those that correlate with perception. Knowledge about a subject's current percept of an ambiguous stimulus is critical for the use of ambiguous stimuli. While humans can be verbally instructed to report their percept, this cannot be done for subjects without language, such as prelingual children and nonhuman animals. However, studies on such subjects are valuable for comparative purposes, and particularly studies on animals are indispensable for unraveling the neuronal underpinnings of the perception of ambiguous and unambiguous stimuli, particularly at the level of single neurons (Bee and Klump, 2004; Fishman et al., 2001; Micheyl et al., 2005). Therefore, there is a strong need for reliable methods of assessing the perception of ambiguous stimuli in nonhuman animals.

To our knowledge, attempts to gain insight into an animal's percept of ambiguous stimuli have been limited to the visual modality, e.g., with stimuli inducing binocular rivalry (Miezin et al., 1981; Logothetis and Schall, 1990; Leopold and Logothetis, 1996) or with stimuli inducing generalized flash suppression (Wilke et al., 2006). In behavioral studies on animals, subjects distinguish different stimuli by exhibiting different motor behaviors. This approach, however, is less effective for an ambiguous stimulus, where two behaviors are possible for the same stimulus. Thus, it is

* Corresponding author. Special Lab Primate Neurobiology, Leibniz Institute for Neurobiology, Brenneckestraße 6, 39118 Magdeburg, Germany. Tel.: +49 391 62 63 94471; fax: +49 391 62 63 95482.
E-mail address: Elena.Selezneva@lin-magdeburg.de (E. Selezneva).

0378-5955/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.heares.2012.08.014

not clear how reinforcement should be appropriately administered, to provide useful and unequivocal feedback when an ambiguous stimulus is probed. In the cited studies, this problem has been addressed by checking response consistency in catch trials with clearly distinguishable stimuli, by comparing behavioral responses to ambiguous stimuli of animals with those of humans, or by modifying the stimuli and observing that behavioral responses change. Here we present an alternative behavioral procedure that avoids the problem of requiring different behaviors for different percepts of an ambiguous stimulus.

In audition, a simple ambiguous stimulus is a repeating sequence of a low-frequency tone A and a high-frequency tone B, presented either in ABAB or ABA_design. This and similar sequences have been widely used in research on auditory scene analysis in humans (van Noorden, 1975; Bregman, 1990; Pressnitzer and Hupe, 2006; Denham and Winkler, 2006; Moore and Gockel, 2012). Depending on the presentation rate of the tones and their frequency separation, listeners perceive the sequence either as one stream, or as two streams, or their perception switches between the two interpretations, i.e., it is bistable or ambiguous in the latter case.

There is converging evidence that non-human animals also perceptually group or organize sequential tones (Hulse et al., 1997; MacDougall-Shackleton et al., 1998; Fay, 1998; Benney and Braaten, 2000; Izumi, 2002). Sequence parameters determining whether tones are perceived as belonging to one stream or to two streams appear to be similar to those for humans. However, there has been no evidence that non-human animals experience bistable auditory percepts, i.e., alternations between one-stream and two-stream percepts.

Based on a pilot experiment on monkeys (see Rahne et al., 2008), some of the authors have recently proposed a new method for assessing the perceptual organization of tone sequences that utilizes response times (RT) to an audio-visual (AV) sequence (see Fig. 1). In this experiment, monkeys were trained to detect the termination of a sequence of light flashes, which was either presented alone, or synchronized in different ways with a sequence of tones that alternated between a low A-tone and a high B-tone. When through many repetitions every third tone was synchronized with a flash, and when the tone sequence stopped shortly before the flashing stopped, the monkeys responded earlier to the termination of the flashes than they did for purely visual sequences; this did not occur when the tone sequence terminated with the flash sequence. RT decreases were maximal when the frequency separation was small (same frequency), and minimal when the frequency separation was intermediate or large (0.5 or 2 octaves). We concluded that RT decreases were related to the perceptual organization of the tone sequences and not only to the frequency separation of the tones (because there were no RT differences between intermediate and large frequency separation). We speculated that the RT decreases were caused by cross-modal facilitation induced by the prematurely terminated tone sequence. The strength of this cross-modal facilitation depended on the match between the perceptual organization of the tone sequence and that of the flash sequence. For both large and intermediate frequency separations, in which the tones would be segregated into two streams, the flashes alternated between being synchronized with the low or high frequency streams. For small frequency separations, in which the tones would be integrated into one stream, the flashes were always synchronized with the same stream. The results suggest that there was less coupling between the visual and auditory sequences for large and intermediate frequency separations than for small frequency separations. We also concluded that the RT decreases can be used to identify perceptually ambiguous tone sequences. However, no evidence for any perceptually ambiguous sequences was obtained in these experiments.

The purpose of the present study was to find tone sequences that are perceptually ambiguous for nonhuman primates. To this end, we again used AV sequences in which a flash sequence was synchronized to a tone sequence and measured RTs, but we performed more comprehensive behavioral testing on monkeys than in the previous experiment (Rahne et al., 2008). This included the use of sequences with more rates and with more refined frequency separations as well as AV sequences in which either every second or third tone was synchronized with a flash. Since, to our knowledge, this paradigm has not yet been used in humans, and to compare RTs with verbal reports on the temporal organization of tone sequences, we complemented our study by performing psychophysical tests on human subjects, using the same AV sequences.

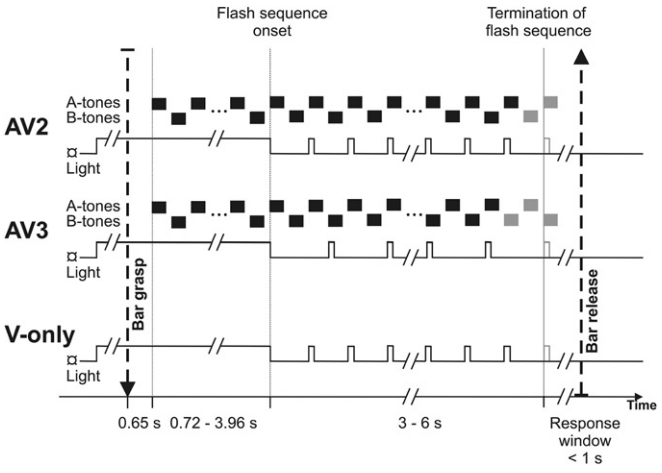

Fig. 1. Experimental paradigm. Audiovisual sequences (AV2 and AV3) and sequences of flashes only (V-only) were used. The sequences were composed of light flashes (solid line) and tones (black rectangles) which alternated between high A-tones and low B-tones. The tone sequence stopped before the flash sequence. Gray rectangles refer to omitted tones; gray lines refer to omitted flashes. Durations of different phases of the behavioral paradigm are indicated underneath the graphic. See Methods for further details.

2. Material and methods

2.1. Subjects

Psychoacoustic tests on animals were approved by the authority for animal care and ethics of the federal state of Saxony Anhalt (No. 43.2-42502/2-802 IfN) and conformed to the rules for animal experimentation of the European Communities Council Directive (86/609/EEC). Experiments were performed on two adult male macaque monkeys (*Macaca fascicularis*), both 6 years old at the beginning of the experiments. Monkey W participated in the pilot study (Rahne et al., 2008), whilst monkey E had previously participated in an auditory detection task. Throughout the experiments, the two monkeys were housed together in a cage in which they had free access to dry food including pellets, bread, corn flakes, and nuts. They earned a large proportion of their water ration during the training sessions, and received the remainder in the form of fresh fruit after each session. On days without behavioral testing they received water and fruit. The body weight and the general appearance of the monkeys were assessed daily.

Psychoacoustic measurements were also performed on 11 human subjects (six male, five female), aged between 20 and 52 years. Four of them had knowledge on the purpose of the study (four of the authors). All of the subjects reported normal hearing

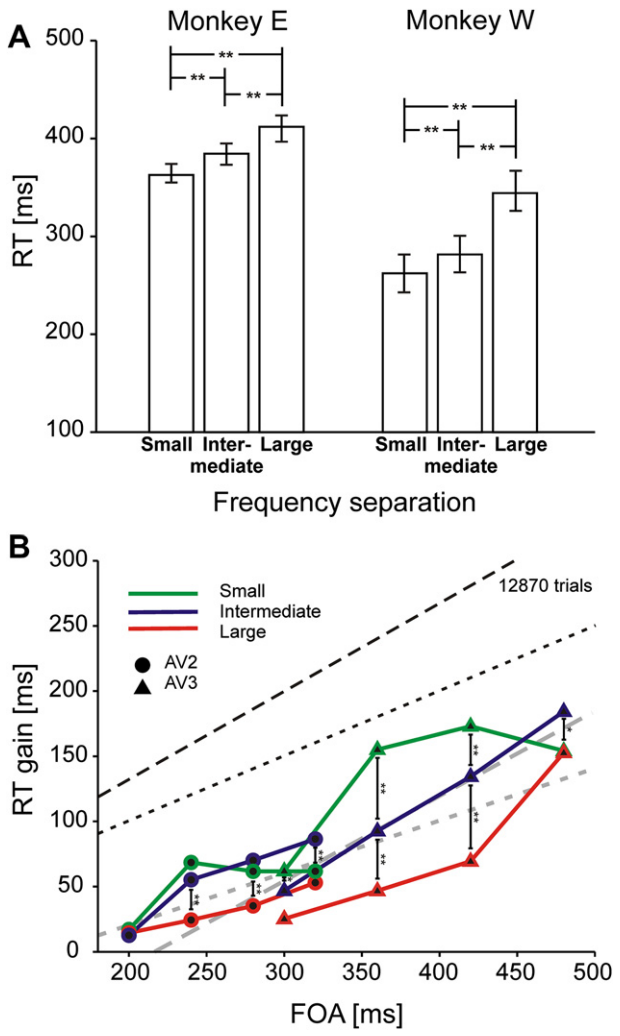

Fig. 2. Auditory stimuli affect visual reaction times of monkeys. (A) RT required to detect the termination of the flash sequence differ between AV3 sequences consisting of tones with small, intermediate and large frequency separation for two monkeys. For each frequency separation data with different flash onset asynchronies were pooled. (B) Response time gains for AV sequences in which the tone sequence terminated before the visual sequence. The plot shows how RT gains depend on flash onset asynchrony (FOA), frequency separation (small [1.5 semitones], green line; intermediate [4.5 semitones], blue line; large [10.5 semitones], red line), and type of AV coupling (circles: AV2 sequences; triangles: AV3 sequences). Statistically significant differences between RT gains measured for different frequency separations are indicated by '*' ($p < 0.05$) or '**' ($p < 0.01$). Dotted and dashed gray lines denote linear regression curves for RT gains of AV2 and AV3 sequences, disregarding frequency separation. RT gains that are expected if subjects responded solely to the termination of the tone sequence are indicated by the dotted (AV2) and the dashed (AV3) black lines.

and no history of hearing disorders. The human psychophysical experiments were conducted in accord with local ethics.

2.2. Apparatus

All experiments were carried out in a double-walled sound-proof room (IAC 1202-A) which was illuminated by three halogen room lights. The monkeys sat in a primate chair whose front compartment accommodated a green light-emitting diode (2 visual degree diameter; 25 cm from the animal), a touch bar, and a water spout (see also Brosch et al., 2004). The water spout was connected through a plastic tube to a magnetic valve which was located outside the sound-proof room. Human subjects sat comfortably on an office chair, watched the LED from a distance of ~40 cm and could touch a button of a response box (TDT RBOX4) located in front of them.

Auditory stimuli were generated on a computer and interfaced with an array processor (Tucker-Davis Technologies, Gainesville), at a sampling rate of 100 kHz. The signal was D/A converted, amplified (Pioneer, A202) and fed to a free-field loudspeaker (Manger, Mellrichstadt), which was placed ~1.2 m and 40° from the midline into the right side of the subjects. The sound pressure level (SPL) was measured with a free field 1/2 inch microphone (40AC, G.R.A.S., Vedbak), located close to the subject's head, connected to a spectrum analyzer (SA 77, Rion). The computer used to generate the auditory and visual stimuli also monitored, controlled, and recorded relevant events of the behavioral procedure.

2.3. Behavioral procedure

The stimuli are illustrated in Fig. 1. A tone sequence that alternated periodically between frequencies A and B (black rectangles) was synchronized with a sequence of flashes (solid line). The time a subject needed to respond to the termination of the flash sequence, i.e., to the first omitted flash (gray line) was measured.

A trial started with the illumination of the LED, which was the signal for the subjects to make contact with the touch bar within 3 s. After maintaining such contact for 650 ms, the tone sequence was turned on. After a variable period of 720–3960 ms the LED was briefly turned off and then started to flash, marking the onset of the AV sequence. Its duration varied between 3000 and 6000 ms. Thus, the tone sequence lasted between 3720 and 9960 ms, which is longer than the time within which 'build-up' usually takes place (e.g., Micheyl et al., 2007). The subjects' task was to release the touch bar as soon as possible after the first omitted flash, irrespective of the presence of the tones. If they did this within 1000 ms, the RT was stored, the trial was scored correct and the monkeys received a water reward (no feedback was given to human participants). The next trial started after a 5-s intertrial interval. A trial was scored incorrect and no reward was administered for bar releases outside the 1000-ms response window. The use of variable sequence durations, together with relatively short behavioral response windows, precluded subjects from simply basing their decisions on estimating the time that had elapsed after trial onset.

All flashes had a duration of 35 ms and all tone bursts had a duration of 80 ms, including 5-ms rise/fall time. Flashes were turned on 9 ms before the corresponding tone, to account for different processing speeds of the visual and auditory modalities (Vroomen and Keetels, 2010). Temporal parameters were chosen to achieve the subjective impression of simultaneous audiovisual events. The intensity of the tone bursts used throughout the experiment was maintained at ~60 dB SPL.

Because it was expected that RTs would depend critically on how the flash sequence was paired with the tone sequence, and on how the tone sequences were perceptually organized, with the magnitude of RT decreases reflecting how frequently a tone sequence would be perceived as a single auditory stream, we systematically tested the effects of the following manipulations of the AV sequences. The selection of parameters was guided by attempting to create as many presumably ambiguous tone sequences as possible.

(1) *AV coupling*: Visual stimuli can accentuate (stress) selected tones and, thus, separate them from other tones, which in turn can change the perceptual organization of the tone sequence, compared to when there are no visual stimuli (Handel, 2006). We made use of these visually induced perceptual changes by combining the flash and the tone sequences in two ways.

Firstly, we synchronized the flashes to every third tone, i.e., the flashes periodically alternated between being synchronized with an A tone and with a B tone, and in this way promoted a temporal organization of the sequence into a periodic alternation between one accented tone and two unaccented tones. This was termed the AV3 sequence.

Secondly, we synchronized the flashes to every second tone, e.g., to the A tones only, and thus promoted a temporal organization of the tone sequence into a periodic alternation between one accented A tone and one unaccented B tone. This was termed the AV2 sequence. Although for this type of AV coupling, RTs were not found to reflect the temporal organization for the small set of sequence parameters tested in the pilot study (Rahne et al., 2008), we nevertheless tested AV2 sequences with other sequence parameters to get insights into the interactions between the auditory and the visual stimuli that underlie RT changes. Conditions in which every flash was synchronized with a tone were not included into the study design because human subjects report feeling uncomfortable with the resulting high flash rates. All AV sequences were compared to visual-only sequences (V-only sequences).

(2) *Sequence rate*: As the temporal organization of a tone sequence depends on its presentation rate (van Noorden, 1975), we presented tone sequences at four rates, specified as the corresponding tone onset asynchrony (TOA). To test mostly ambiguous sequences, we used TOAs of 100, 120, 140, and 160 ms. Resulting flash presentation rates, expressed as flash onset asynchronies (FOA), were 200, 240, 280, and 320 ms for AV2 sequences and 300, 360, 420, and 480 ms for AV3 sequences. For V-only sequences we used all 8 FOAs. Because of the variable sequence durations, the number of tones and flashes also varied considerably from trial to trial. For a TOA of 100 ms, for example, the initial part of the AV sequence consisted of 10–40 tones. The subsequent bimodal part consisted of 19–60 tones and 10 to 30 flashes for AV2 sequences and 19–58 tones and 7 to 20 flashes for AV3 sequences.

(3) *Frequency separation*: As the temporal organization of a tone sequence depends on the frequency separation of the A and B tones (van Noorden, 1975), three frequency separations were tested that are known to result in different percentages of reports of segregated percepts. To test mostly ambiguous sequences, we used 0.1243 octaves (1.5 semitones; 'small'), 0.3741 octaves (4.5 semitones; 'intermediate'), and 0.8750 octaves (10.5 semitones; 'large'). We used 1000 and 1090 Hz for condition small, 1000 and 1296 Hz for condition intermediate, and 1000 and 1834 Hz for condition large.

Our pilot study (Rahne et al., 2008) showed that RTs depended on the frequency separation, and reflected the perceptual organization of the tone sequence, only if the tone sequence terminated shortly before the flash sequence. Hence, we also used sequences with asynchronous terminations in which the tone sequence ended before the flash sequence. For AV2 sequences, the tone sequence ended one tone before the flash sequence. For AV3 sequences, the tone sequence ended two tones before the flash sequence. To reduce responses to the termination of the tone sequence, 5–30% of the trials were catch trials in which the tones continued for 1000 ms after the flashing was stopped (this corresponded to the end of the response time window, see above).

During individual sessions, monkeys were tested with one frequency separation, one TOA (and thus two FOAs), the three types of AV couplings (AV2, AV3, and V-only), the full range of sequence durations, and sequences in which either the tones or the flashes were stopped first. For humans, one TOA with all three frequency separations was tested in a single session.

All RTs were measured relative to the 'onset' of the first omitted flash. To allow for the variation of RTs across behavioral sessions, for each session we computed a 'RT gain'. This was defined as the difference between the median RT to all V-only sequences presented during a session, and the RT for a specific AV sequence. It should be noted that RTs did not systematically depend on the flash rate for V-only sequences (Fig. 3). Thus, RTs obtained with V-only sequences presented at different FOA could be combined. RT gains from different sessions and from the two monkeys were thus combined to give median RTs, separately for all AV conditions. Results for AV2 sequences in which the flashes were synchronized with the A tones only were similar to those for sequences in which the flashes were synchronized with the B tones only. Therefore, data for the two types of AV sequence were combined. Although RTs varied slightly with sequence duration, and RT differences were maximal at intermediate durations, these variations were similar for all frequency separations. Thus, RTs obtained for different sequence durations were combined. We used Kruskal–Wallis tests to compare samples from two or more groups of different AV conditions, with a significance level of $p \leq 0.05$.

2.4. Training procedure for monkeys

We first trained the two monkeys to perform a task with visual stimuli only. They learned to grasp and hold the touch bar after the light was turned on, to keep on holding it during the flashing, and only to release the touch bar after the flashing had stopped. This was done for 21 sessions (15500 trials) for monkey W and 11 sessions (4000 trials) for monkey E, at the end of which they scored correctly in >84% of the trials. We then added a tone sequence to the flash sequence. During the following 12 sessions for monkey W and 29 sessions for monkey E, we observed that performance improved to ~94% and RTs slowly decreased. The results presented here are from subsequent sessions during which no further RT decreases occurred, i.e., 27 sessions (5486 trials) for monkey W and 48 sessions (7384 trials) for monkey E. The data for monkey E were obtained from the sessions directly following the 29 sessions during which AV sequences were introduced. Monkey W first performed 21 sessions with other sequence parameters after the 12 sessions during which AV sequences were introduced. Because of the different parameters, these data were not considered here, except for the visual-only condition.

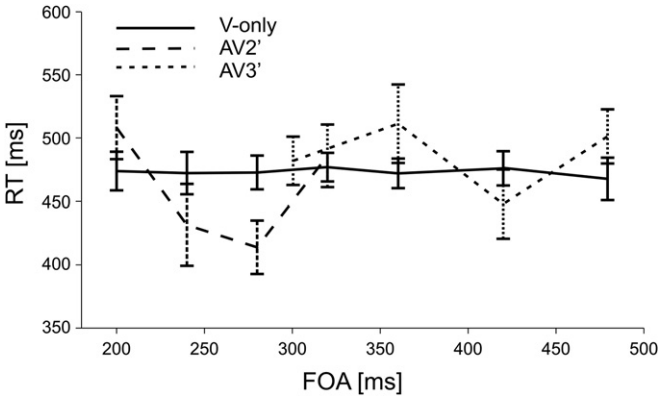

Fig. 3. Response times for visual-only sequences and for AV sequences in which the tone sequence terminated after the visual sequence. The plots shows the medians for different sequences and for different FOA.

3. Results

3.1. Monkey experiment

We confirmed and extended the observations of our pilot study (Rahne et al., 2008) that monkeys responded earlier to the termination of the flash sequence when it was synchronized with a tone sequence and when this tone sequence terminated shortly before the flash sequence (Fig. 2). We also confirmed that response times (RTs) depended on the frequency separation between the tones.

3.1.1. Reaction time depends on termination asynchrony

The main factor influencing RT changes was the termination asynchrony between the tone and flash sequences, defined as the time between the first omitted tone and the first omitted flash. The changes of RTs were expressed as RT gain. When RTs measured for AV sequences with different frequency separations but with the same TOA were combined, the RT gain increased with FOA (Fig. 2B, gray lines). The changes in RT for asynchronously terminated AV sequences were not due to changes in FOA, because for V-only sequences, RTs did not vary with FOAs (Fig. 3, Friedman's test, $p = 0.64$). The dependence of RT gain on FOA could be modeled well with a linear regression (AV3: $R^2 = 0.63$, $F = 16.9$, $p = 0.0021$, $df = 11$; AV2: $R^2 = 0.59$, $F = 14.5$, $p = 0.0034$, $df = 11$) although with different slopes for AV2 and AV3 sequences (0.414 and 0.655, respectively). The two slopes, however, were more similar to each other when RT gain was expressed as a function of termination asynchrony (0.817 and 0.941, respectively). The growth of termination asynchrony with FOA is indicated by the blue and red dotted lines in Fig. 2B. Because the ratio of the two slopes were more similar to each other than the corresponding slope ratios observed for the FOA dependence, our findings strongly suggest that, for both types of AV couplings, RT gain depended mostly on the size of termination asynchrony.

No systematic changes of RTs were observed when the tone sequence continued after the flash sequence had been stopped, even though significant differences were present at individual FOAs (Fig. 3, broken lines). The invariance of RTs was seen for a wide range of flash rates, i.e., for FOAs between 200 and 480 ms, and for both AV2 and AV3 sequences and indicates that the termination of the tone sequence was not sufficient to elicit responses from the monkeys.

Although RT gain increased with termination asynchrony, absolute sizes of RT gains were always smaller than the termination asynchrony (note that the two black lines were always above the two gray regression lines). This indicates that the monkeys' behavior was not fully controlled by the tone sequence. Interestingly, for FOAs around 300 ms, RT gains were similar or even smaller for AV3 sequences than for AV2 sequences, even though the termination asynchrony was larger for AV3 sequences and two tones, and for the former two tones had been omitted before the first omitted flash. This suggests different facilitatory influences of the termination of the tone sequence for AV2 and AV3 sequences.

3.1.2. Reaction time depends on frequency separation

Further insights into how the tone sequence affected the detection of the flash sequence termination can be gained from the observation that RTs depend on the frequency separation of the tones. Fig. 2 shows that, for both monkeys, RT gains increased with decreasing frequency separation of the tones. Frequency effects were present at most FOAs, and were more pronounced as well as more consistent for AV3 than for AV2 sequences. This might be partially due to the larger termination asynchronies that were used for the AV3 sequences. AV3 sequences presented at intermediate FOAs (360 and 420 ms) yielded significantly different RT gains for the frequency separations of small, intermediate, and large (pairwise U-tests, each $p < 0.01$). At the smallest FOA (300 ms) as well as at the largest FOA (480 ms), RT gains differed significantly only between sequences with small and intermediate frequency separations (U-test, $p < 0.05$). For AV2 sequences, by contrast, significantly different RT gains were found only between intermediate and large frequency separations. This suggests different facilitatory influences of the frequency separation between the tones for AV2 and AV3 sequences.

3.1.3. Reaction time reflects perceptual organization

We performed a receiver operating characteristic (ROC) analysis on RT gains (Britten et al., 1992; Green and Swets, 1966) to estimate the fraction of trials in which the tone sequence with an intermediate frequency separation was perceived as a single auditory stream or as two streams. To achieve this, we compared the RT gain distribution for an AV3 sequence that is mostly perceived as one stream (i.e., a sequence with a small frequency separation), with the distribution for an AV3 sequence that is mostly perceived as two streams (i.e., a sequence with a large frequency separation). From this, we calculated how frequently different RT gains would be classified as being drawn from either distribution (Fig. 4A, B). Fig. 4C shows resulting ROC curves obtained from all trials performed by the two monkeys for each of the four TOAs tested. To find the RT gain that best discriminated small from large frequency separation, we searched for the point on the ROC curve that was closest to the upper left coordinate of the ROC plane. For a TOA of 120 ms, this point was at a RT gain of 100 ms. Thus this RT gain distinguished best the RT distributions obtained for sequences with large and small frequency separations. Applying the same criterion for sequences with an intermediate frequency separation suggested that this sequence was perceived in 49% of the trials as two streams and in the remaining trials as one stream. For a TOA of 140 ms, best discrimination was at a RT gain of 130 ms. This suggests that AV sequences with an intermediate frequency separation were perceived in 52% of the trials as two streams and in 48% of the trials as one stream. Consequently, these configurations of streams seemed to be perceived as ambiguous. At the largest and smallest TOAs, the RT distributions were virtually identical. Thus no such RT gains for best separation were determined.

Another argument that AV sequences with an intermediate frequency separation are perceptually ambiguous arises from the shape of RT distributions, which, particularly at a TOA of 120 ms, were bimodal. This may reflect the superposition of two RT distributions, one for AV sequences that are perceived as one stream and another for AV sequences that are perceived as two streams.

3.2. Comparison of monkey with human data

Four human subjects (authors) were tested under similar conditions as the monkeys, with the same asynchronously terminated AV sequences and V-only sequences (Fig. 1).

Fig. 5A shows that the human subjects exhibited dependences of RT gain on flash rate, frequency separation, and AV coupling that were similar to those observed for the monkeys (Fig. 2B). RT gain increased with FOA (i.e., decreased with flash rate), and for AV3 sequences, also with frequency separation. As for the monkeys, RT gain increased with, and was always smaller than, the termination asynchrony. This indicates that the termination of the tone sequence did not trigger a response, but rather enabled the subjects to respond faster to the termination of the flash sequence. Compared to the monkeys, the human subjects generally gained more from the presence of the concurrent tone sequence, particularly for tone sequences with large frequency separations.

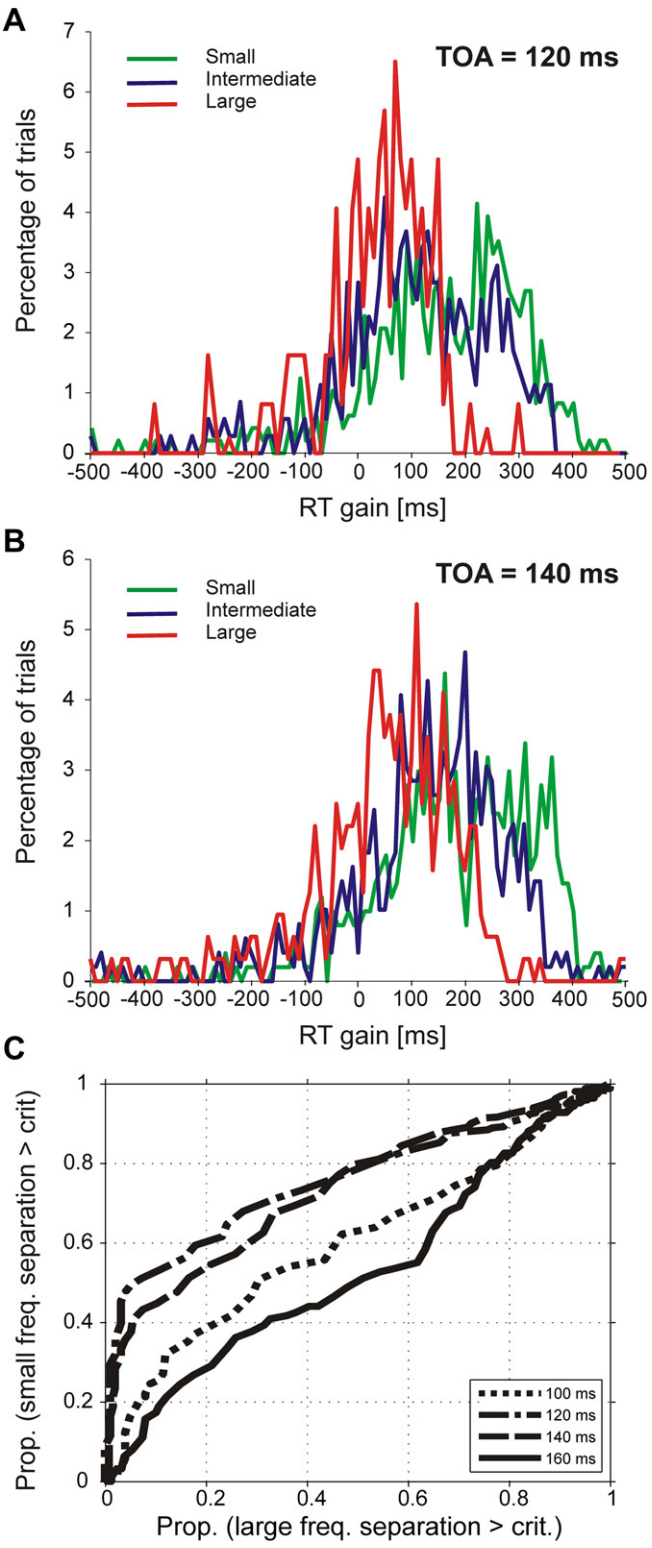

Fig. 4. Determination of response time gains that best discriminate trials in which an AV sequence is perceived as one stream from trials in which an AV sequence is perceived as two streams. (A): Distributions of response time (RT) gains for AV3 sequences with different frequency separations (small, intermediate, and large) obtained at a TOA of 120 ms. (B) Corresponding distributions for a TOA of 140 ms. (C) ROC-plot. Fraction of trials in which a specific RT gain criterion belongs to the RT distribution with a large frequency separation versus that it belongs to the RT distribution with a small frequency separation. For each TOA, the diamond-shaped markers correspond (from left to right) to RT gains of 0, 100, 200 and 300 ms. Note that at TOA = 120 ms, maximal d' was 1.80, and at TOA = 140 ms maximal d' was 1.55. For the smallest (100 ms) and the largest (160 ms) TOA, d' was <1.

In sixteen additional sessions, we assessed how the same subjects perceptually organized the tones in the AV sequences. Subjects were exposed to AV2, AV3 and auditory-only sequences and were instructed to attend both to the tones and the flashes and to indicate when they perceived two auditory streams by pressing a button. All other periods were considered as periods in which the subjects perceived one stream.

In accordance with previous studies conducted using purely auditory sequences (van Noorden, 1975; Bregman, 1990), the percentage of cases for which subjects reported hearing two segregated streams increased with increasing frequency separation of the tones (Fig. 5B, solid lines). This was observed for all tone presentation rates tested (TOA between 100 and 160 ms). For several frequency separations, there was a trend that the more rapidly presented sequences were more frequently segregated into two streams. The perceptual organization of the tone sequence changed when it was synchronized with a flash sequence, demonstrating an influence of visual stimuli on auditory stimuli (Fig. 5B; broken lines). Generally, AV2 coupling had a segregating effect on the perception of the tone sequence, while AV3 coupling had an integrating effect. This phenomenon was seen at most TOAs and for most frequency separations.

The two experiments on humans enabled us to directly compare RT gains with reports on the temporal organization of the AV sequences. Fig. 5C confirms that RT gain was greater the more often the sequence was perceived as one stream ($r = 0.5$, $p < 0.05$, one-tailed Spearman correlation). This suggests that, in humans, the RTs required to detect the termination of a flash sequence closely reflect the temporal relationship of tone sequences and flash sequences.

To rule out the possibility that the results for RT gain (Fig. 5A) were affected by knowledge the subjects had on the purpose of the study, we repeated the RT experiment using seven additional subjects without such knowledge. To further reduce the probability that they developed this knowledge over the course of the experiment, each of the subjects performed only one session with one TOA (120 ms). Fig. 5D shows that these uninformed subjects exhibited a qualitatively similar dependence of RT gain on frequency separation as the informed subjects and the highly-trained monkeys.

4. Discussion

The present study provides experimental evidence that tone sequences can be perceptually ambiguous for macaque monkeys. This was shown by synchronizing a sequence of periodically alternated tones with a flash sequence and measuring the time the monkeys required to respond to the termination of the flash sequence.

We found that compared to a visual-only sequence, the monkeys responded more quickly when the tone sequence terminated shortly before the flash sequence, and that this RT gain depended on the frequency separation of the tones. RT gain was largest when the frequency separation was so small that the sequential tones were integrated into one auditory stream. RT gain was smallest when the frequency separation was so large that the sequential tones were segregated into two streams. RT gain was moderate for intermediate frequency separations. In the following, we argue that the latter tone sequences are perceptually ambiguous for monkeys, i.e., switch between being perceived as one stream or as two streams. Further, we argue that RT gain is chiefly caused by cross-modal facilitation induced by the first omitted tone in the AV sequence. The strength of this cross-modal facilitation depends on the match between the perceptual organization of the tone sequence and that of the flash sequence.

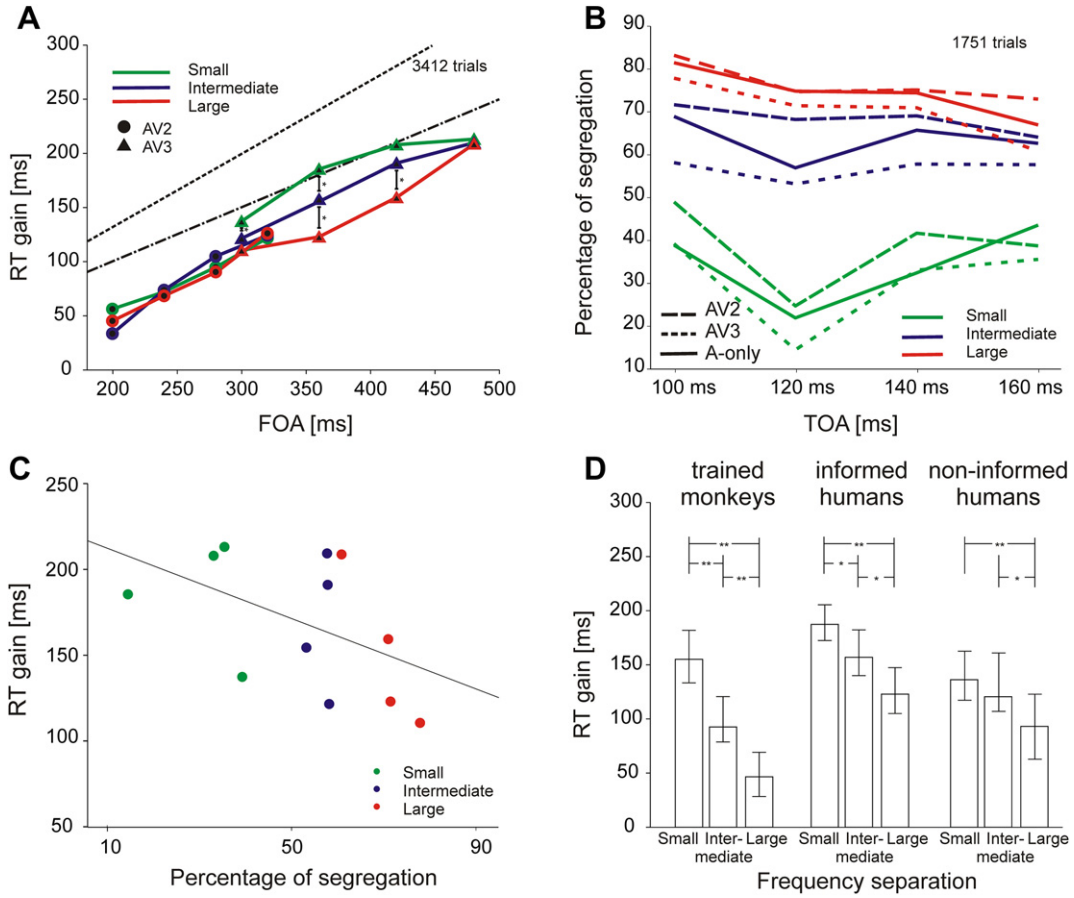

Fig. 5. Results of the experiments performed on humans. (A) RT gains for different FOA, frequency separations, and AV couplings (conventions as in Fig. 2B). Dashed (AV2) and the dotted (AV3) black lines represent RT gains those are expected if subjects responded solely to the termination of the tone sequence. Note that the results were similar to those obtained in monkeys. (B) Auditory perceptual organization of AV2, AV3 and auditory-only (A-only) sequences for different tone onset asynchronies (TOA) and frequency separations. Percentage of segregation reflects the total time in which subjects reported hearing two streams, divided by the total time of sequence presentation. (C) Relationship between auditory streaming organization and RT gain in humans for the AV3 data displayed in panels A and B. Black line shows linear regression. (D) Comparison of the frequency dependence of RT gains at a TOA of 120 ms between two monkey subjects (compare Fig. 2) and human participants with ('informed', $n = 4$) and without ('non-informed'; $n = 7$) knowledge about the purpose of the study.

4.1. Response time gain depends on termination asynchrony

The most important sequence property for RT gain to occur was the termination asynchrony. Only when the tone sequence terminated shortly before the flash sequence was the RT decreased (Fig. 2). RT was unchanged when the tone and flash sequence stopped at the same time, or when the tone sequence continued after the flash sequence had ended (Fig. 3; Rahne et al., 2008). This result was found for both types of AV couplings tested here (AV2 and AV3); thus synchronizing visual with auditory stimuli will not necessarily result in a bimodal processing advantage (Spence, 2007).

Although the behavior of the monkeys was affected by the termination of the tone sequence, this termination affected only the timing of motor responses but was not responsible for eliciting such responses. This was reflected by similar percentages of correct responses for the different types of AV sequences tested here and previously (Rahne et al., 2008). Hence, the decision as for whether to respond was solely controlled by the termination of the flash sequence.

Similar cross-modal facilitation occurs when stimuli of different modalities (auditory, visual, tactile and smell) are presented congruently in closed spatial, temporal or semantic proximity. In addition to affecting RTs (Sakata et al., 2004; Foxton et al., 2010; Merlo et al., 2010), cross-modal facilitation also improves accuracy (Sakata et al., 2004; Rossi et al., 2008; Merlo et al., 2010) and perceptual sensitivity (McDonald et al., 2000; Frassinetti et al., 2002; Bolognini et al., 2005).

For AV2 and particularly for AV3 sequences, RT gain increased with increasing termination asynchrony (Fig. 2B). This was found for termination asynchronies between 100 ms and 320 ms and could not be explained by differences in FOA (Fig. 3). Unfortunately, we were not able to test other termination asynchronies extensively because our experiment primarily focused on the influence of the frequency of the tones. Nevertheless, it seems that cross-modal facilitation has an inverted U-shaped dependence on termination asynchrony. This is suggested by two findings: (1) no RT gain occurred for synchronously terminated AV sequences, i.e., when the first omitted flash and the first omitted tone coincided; (2) no RT gain occurred for AV sequences in which the tone sequence stopped after the flashes (Rahne et al., 2008). The AV coupling may give rise to neuronal entrainment (Schröder and Lakatos, 2009) reflected in phase resets, which may slowly decay after one input stream is terminated.

4.2. Relationship between perceptual interactions between elements of the AV sequence and response times

Two further observations suggest that RT gain depends on additional parameters of the AV sequence. Firstly, RT gain was

larger for AV2 sequences than for AV3 sequences. This is most obvious around a FOA of 300 ms, where the termination asynchronies of the two types of AV couplings differed by only 50 ms but they produced similar RT gains (Fig. 2B). This suggests that, for a given termination asynchrony, there was more cross-modal facilitation for AV2 sequences than for AV3 sequences and, more generally, that the strength of cross-modal facilitation depends on how the tone and flash sequences are coupled. A possible reason for different cross-modal facilitation is that in AV3 sequences two tones were omitted before the first omitted flash, so the second omitted tone contributed little, if at all, to cross-modal facilitation and thus to RT gain. Also, cross-modal facilitation could depend on the fraction of tones that are synchronized with a flash. This was higher for AV2 sequences, which could have resulted in stronger perceptual binding of the flashes with the tones.

The second observation suggesting that RT gain depends on additional parameters is that a clear frequency dependence of RT gains was only found for AV3 sequences. This suggests a role of the composition of AV sequences for the cross-modal facilitation induced by the first omitted tone, and thus a role of the strength of perceptual binding between the elements of the AV sequence. The binding could, as a first hypothesis, exclusively rely on interactions between the tones, which selectively affect how well tone omission can be detected during the auditory temporal grouping, i.e., it could be related to whether or not the last presented and the first omitted tone are from the same stream or from different streams. As a second hypothesis, cross-modal facilitation could rely on the strength of perceptual binding between the tone sequence and the flash sequence, i.e., the match of the perceptual organizations of the two sequences. In the following we argue that the current results are more consistent with the second hypothesis.

Fig. 6 schematizes possible perceptual relationships between the elements of AV sequences. The left column depicts the perceptual organizations of the tone sequences that emerge from the auditory Gestalt principle of frequency similarity. Sequences with small frequency separations are integrated into a single stream (panel A for AV3 sequences, panel C for AV2 sequences). Sequences with large separations split into two segregated streams, one consisting of the a-tones and the other of the b-tones (panels B and D for AV3 and AV2 sequences).

Starting with the first hypothesis that cross-modal facilitation is mainly due to auditory interactions, we expect strong cross-modal facilitation and thus a large RT gain when all tones are integrated into a single stream (panels A, C; indicated by the solid straight arrow). In contrast, we expect little cross-modal facilitation and thus a small RT gain when the tones are segregated into two streams (panels B, D; indicated by the dashed straight arrow). Consequently, sequences that are integrated into a single stream should lead to a larger RT gain than sequences that are segregated into two streams. This is inconsistent with our observation of a frequency dependence of RT gain for AV3 sequences only (Fig. 2).

The frequency dependence of RT gain is, however, consistent with the second hypothesis, namely that cross-modal facilitation depends on the relationship between the concurrent auditory and visual streams. In three conditions (panels A, C, D), the flashes are always synchronized with tones perceiving in the same auditory stream. For AV sequences with a small frequency separation (which are integrated into one stream, panels A and C), the flashes are synchronous with every second tone of this auditory stream (i.e., the a-tones). For AV2 sequences with a large frequency separation (which segregate into two streams; panel D), the flashes are synchronous with every tone of one of the two auditory streams (again the a-tones). Thus, there is a match between the perceptual organization of the auditory and the visual sequences in these three conditions, which could result in strong coupling between the tones and the flashes and thus strong cross-modal facilitation and

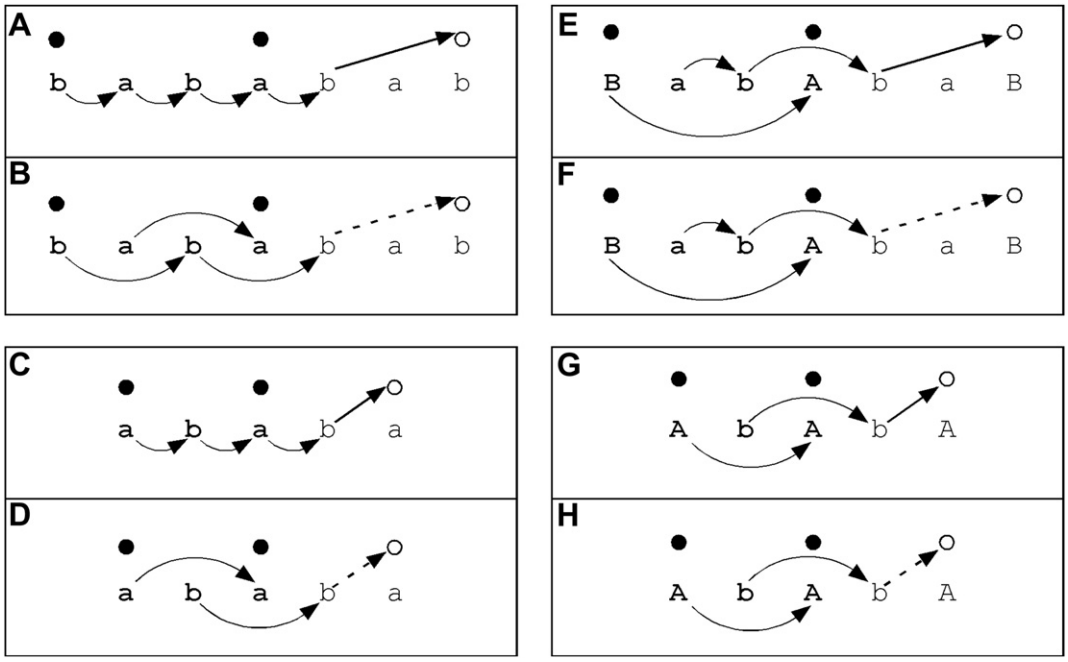

Fig. 6. Schema of perceptual relationships between elements of an AV sequence and their contribution to response time gain. Left column: The perceptual organizations of the tone sequence due to the Gestalt principle of frequency similarity, when the tones are integrated into a single stream (A, C) or when they are segregated into two streams (B, D). (A) and (B) show organizations for AV3 sequences, and (C) and (D) for AV2 sequences. Presented tones are indicated by bold letters, omitted tones by plain letters. The temporal evolution of the streams is indicated by curved arrows. The straight arrow indicates cross-modal facilitation induced by the tone omission. Right column: The perceptual organizations of the tone sequence that emerge from the accenting effect of the flashes onto selected tones. (E, F) for AV3 sequences, the tone sequence splits into two streams, which alternate between one accented tone and two unaccented tones (indicated by uppercase and lowercase letters). (G, H) for AV2 sequences, the tone sequence also splits into two streams, which alternate between one accented tone and one unaccented tone.

large RT gains. In contrast, for AV3 sequences with a large frequency separation, the flash sequence alternates between being synchronized with tones perceived in the first and the second streams (panel B). Thus, there is a mismatch between the perceptual organization of the auditory and visual sequences, resulting in weak audiovisual coupling and thus little cross-modal facilitation and small RT gains. Consequently, AV3 sequences with small and large frequency separations should have different RT gains, whereas AV2 sequences with small and large frequency separations should have similar RT gains. This is in accordance with the results shown in Fig. 2B.

The differential frequency dependence of AV2 and AV3 sequences also suggests that RT gain is unlikely controlled by influences of the visual stimuli on the auditory stimuli, i.e., by the accentuation of specific tones by the flashes. This can be deduced from the right column of Fig. 6. Panels E–H depict the perceptual organizations of the tone sequences that might emerge from the accenting effect of the flashes on specific tones, and which might segregate these tones from the remaining tones. For AV3 sequences, the tone sequence splits into two streams, which alternate between one accented tone and two unaccented tones (indicated by uppercase and lowercase letters in panels E and F). For AV2 sequences, the tone sequence also splits into two streams which, however, alternate between an accented tone and an unaccented tone (panels G and H). These panels show that AV sequences with small and large frequency separations have the same relationship between the tones and the flashes and that RTs should be similar for AV2 and AV3 sequences – unlike what is shown in Fig. 2B. While visuo-to-auditory influences cannot explain the current results on RT, such influences have been demonstrated previously (O'Leary and Rhodes, 1984; Rahne et al., 2008; Marozeau et al., 2010) and are even present in one of our experiments on humans (Fig. 5B).

4.3. Response time gains at different presentation rates

Frequency separation had a clear effect on RT gain only for AV3 sequences when the tones were presented at higher rates (TOAs between 100 and 140 ms) but not when they were presented at the lowest rate (TOA = 160 ms). A possible reason for this is that at a TOA of 160 ms, the perceptual organization of the tone sequence is less determined by the physical properties of the sequence and is thus less 'obligatory' and more under 'voluntary' control of the subjects than at higher presentation rates (van Noorden, 1975; Bregman, 1990). Since subjects can decrease their RT when they integrate the tone sequence into a single stream, subjects may be biased towards integration, resulting in similar RTs for all frequency separations. No such biasing can be assumed when human subjects are asked to report their streaming percept.

4.4. Response time gains as an index of the perceptual organization of tone sequences

The discussion thus far leads to the conclusion that RT gain is determined by the perceptual organization of the tone sequences. Small RT gains occur when the tone sequence segregates into two streams. Large RT gains occur when the tone sequence is integrated into a single stream. Because RT gains for small and large frequency separations were significantly different from each other and from the RT gains for an intermediate frequency separation, we conclude that the perceptual organization of AV3 sequences with intermediate frequency separation is different from that of sequences with small or large frequency separations. Because tone sequences are most of the time perceived either as one or as two auditory streams, we conclude that sequences with intermediate frequency separation switch between being perceived as one or as two auditory streams, i.e., they are perceptually ambiguous.

Although RT gains are related to the perceptual organization of a tone sequence, RTs cannot unequivocally reveal the perceptual organization in individual trials. This holds both for perceptually ambiguous and for perceptually unambiguous tone sequences. We found that the RT distribution for sequences with a small frequency separation overlaps with the RT distribution for sequences with a large frequency separation (Fig. 4). Thus RT gain can be used to assess the perceptual organization of a tone sequence only in a probabilistic manner.

Independent support for our conclusion that RT gain can be used to assess the perceptual organization of tone sequences in monkeys comes from the experiments we performed under very similar conditions on informed and uninformed human subjects (Fig. 5). The data from these experiments show that RT gain for humans has a similar dependence on termination asynchrony, frequency separation, and type of AV coupling as for the monkeys. This suggests that, also for humans, RT gains are most likely caused by cross-modal facilitation, the strength of which is determined by the match of the perceptual organization of the tone sequence with that of the flash sequence.

The use of human subjects enabled us to directly compare subjective reports to RT gains for AV sequences. Fig. 5C shows that there was a correlation between the size of the RT gain, and the percentage of trials in which a tone sequence was perceived as two auditory streams. Specifically, intermediate RT gains were frequently associated with sequences which were rated as segregated on about 50% of trials. Because of the similar dependences of RT gains in monkeys and humans, this provides additional support for our interpretation that AV sequences with intermediate frequency separations are perceptually ambiguous for monkeys.

5. Conclusion

The present study on monkeys and humans suggests that use of AV sequences together with measurements of RT gains provides a means for identifying perceptually ambiguous stimuli. The procedure described here can be applied to test psychological models of perceptual organization, to compare perceptual organizations of sequential stimuli in different species, and to identify neuronal mechanisms that underlie the integration and segregation of sequential stimuli.

Acknowledgments

We thank C. Bucks for assistance in animal care, C. Micheyl and A. Brechmann for comments on earlier versions of this manuscript, and J. Lovell for improving the English. Supported by the Deutsche Forschungsgemeinschaft (SFB TR 31, SFB 779).

References

Bee, M.A., Klump, G.M., 2004. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. J. Neurophysiol. 92, 1088–1104.

Benney, K., Braaten, R., 2000. Auditory scene analysis in estrildid finches (*Taeniopygia guttata* and *Lonchura striata* domestica): a species advantage for detection of conspecific song. J. Comp. Psychol. 114, 174–182.

Bolognini, N., Frassinetti, F., Serino, A., Làdavas, E., 2005. "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp. Brain Res. 160, 273–282.

Bregman, A.S., 1990. Auditory Scene Analysis. The Perceptual Organization of Sound. MIT Press, Cambridge, MA.

Britten, K.H., Shadlen, M.N., Newsome, W.T., Movshon, J.A., 1992. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 12, 4745–4765.

Brosch, M., Selezneva, E., Bucks, C., Scheich, H., 2004. Macaque monkeys discriminate pitch relationships. Cognition 91, 259–272.

Denham, S.L., Winkler, I., 2006. The role of predictive models in the formation of auditory streams. J. Physiol. Paris 100, 154–170.
Fay, R., 1998. Auditory stream segregation in goldfish (*Carassius auratus*). Hear. Res. 120, 69–76.
Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hear. Res. 151, 167–187.
Foxton, J.M., Riviere, L.D., Barone, P., 2010. Cross-modal facilitation in speech prosody. Cognition 115, 71–78.
Frassinetti, F., Bolognini, N., Làdavas, E., 2002. Enhancement of visual perception by crossmodal visuo-auditory interaction. Exp. Brain Res. 147, 332–343.
Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. Wiley, New York.
Handel, S., 2006. Perceptual Coherence: Hearing and Seeing. Oxford Univ Press.
Hulse, S., MacDougall-Shackleton, S., Wisniewski, A., 1997. Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). J. Comp. Psychol. 111, 3–13.
Izumi, A., 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113–B122.
Leopold, D., Logothetis, N., 1996. Activity changes in early visual cortex reflect monkeys percepts during binocular rivalry. Nature 379, 549–553.
Logothetis, N., Schall, J., 1990. Binocular motion rivalry in macaque monkeys: eye dominance and tracking eye movements. Vis. Res. 30, 1409–1419.
MacDougall-Shackleton, S., Hulse, S., Gentner, T., White, W., 1998. Auditory scene analysis by European starlings (*Sturnus vulgaris*): perceptual segregation of tone sequences. J. Acoust. Soc. Am. 103, 3581–3587.
Marozeau, J., Innes-Brown, H., Grayden, D.B., Burkitt, A.N., Blamey, P.J., 2010. The effect of visual cues on auditory stream segregation in musicians and non-musicians. PLoS One 5 (6), e11297.
McDonald, J.J., Teder-Sälejärvi, W.A., Hillyard, S.A., 2000. Involuntary orienting to sound improves visual perception. Nature 407, 906–908.
Merlo, J.L., Duley, A.R., Hancock, P.A., 2010. Cross-modal congruency benefits for combined tactile and visual signaling. Am. J. Psychol. 123, 413–424.
Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P., 2005. Perceptual organization of tone sequences in the auditory cortex of awake macaques. Neuron 48, 139–148.
Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E., 2007. The role of auditory cortex in the formation of auditory streams. Hear. Res. 229, 116–131.
Miezin, F., Myerson, J., Julesz, B., Allman, J., 1981. Evoked potentials to dynamic random-dot correlograms in monkey and man: a test for cyclopean perception. Vis. Res. 21, 177–179.
Moore, B.C.J., Gockel, H., 2012. Properties of auditory stream formation. Phil Trans. R Soc. B 367, 919–931.
O'Leary, A., Rhodes, G., 1984. Cross-modal effects on visual and auditory object perception. Percept Psychophys. 35, 565–569.
Pressnitzer, D., Hupé, J.M., 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. Curr. Biol. 16, 1351–1357.
Rahne, T., Deike, S., Selezneva, E., Brosch, M., König, R., Scheich, H., Böckmann, M., Brechmann, A., 2008. A multilevel approach towards neuronal mechanisms of streaming. Brain Res. 1220, 118–131.
Rossi, S., De Capua, A., Pasqualetti, P., Ulivelli, M., Fadiga, L., Falzarano, V., Bartalini, S., Passero, S., Nuti, D., Rossini, P.M., 2008. Distinct olfactory cross-modal effects on the human motor system. PLoS One 3 (2), e1702.
Sakata, S., Yamamori, T., Sakurai, Y., 2004. Behavioral studies of auditory-visual spatial recognition and integration in rats. Exp. Brain Res. 159, 409–417.
Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci. 32, 9–18.
Spence, C., 2007. Audiovisual multisensory integration. Acoust. Sci. Technol. 28, 61–70.
van Noorden, L.P.A.S., 1975. Temporal coherence in the perception of tone sequences. Ph.D. thesis, Eindhoven University of Technology.
Vroomen, J., Keetels, M., 2010. Perception of intersensory synchrony: a tutorial review. Atten. Percept. Psychophys. 72, 871–884.
Wilke, M., Logothetis, N., Leopold, D., 2006. Local field potential reflects perceptual suppression in monkey visual cortex. Proc. Natl. Acad. Sci. USA 103, 17507–17512.