

Time-critical integration of formants for perception of communication calls in mice

Diana B. Geissler and Günter Ehret*

Department of Neurobiology, University of Ulm, D-89069 Ulm, Germany

Edited by Michael M. Merzenich, University of California, San Francisco, CA, and approved April 23, 2002 (received for review November 13, 2001)

Brain mechanisms in humans group together acoustical frequency components both in the spectral and temporal domain, which leads to the perception of auditory objects and of streams of sound events that are of biological and communicative significance. At the perceptual level, behavioral data on mammals that clearly support the presence of common concepts for processing species-specific communication sounds are unavailable. Here, we synthesize 17 models of mouse pup wriggling calls, present them in sequences of four calls to the pups' mothers in a natural communication situation, and record the maternal response behavior. We show that the biological significance of a call sequence depends on grouping together three predominant frequency components (formants) to an acoustic object within a critical time window of about 30-ms lead or lag time of the first formant. Longer lead or lag times significantly reduce the maternal responsiveness. Central inhibition seems to be responsible for setting this time window, which is also found in numerous perceptual studies in humans. Further, a minimum of 100-ms simultaneous presence of the three formants is necessary for occurrence of response behavior. As in humans, onset-time asynchronies of formants and formant durations interact nonlinearly to influence the adequate perception of a stream of sounds. Together, these data point to common rules for time-critical spectral integration, perception of acoustical objects, and auditory streaming (perception of an acoustical Gestalt) in mice and humans.

auditory objects | auditory streaming | formant grouping | time-critical processing | sound communication

Which sound features are important for the perception and discrimination of acoustic patterns has become a central and controversially discussed question in research on auditory psychophysics and perception in humans as explained in detail by Hirsh and Watson in their 1996 review (1). A stream of acoustic patterns requires not only auditory analysis but also grouping together of analyzed spectral and temporal elements according to their coherence and synchrony to be perceived as auditory objects, phonemes, and words in speech or instruments and melodies in music (2–5). The wealth of data on perception of complex sounds in humans contrasts with the lack of knowledge of the neural mechanisms underlying human auditory grouping and object perception. On the other hand, neurophysiological studies in amphibians, birds, and mammals have shown that neurons in higher auditory centers respond preferentially to combinations of spectral and temporal properties of species-specific communication calls (6–14) or of bats' own echolocation sounds (15–19), thus demonstrating neural mechanisms for acoustic object perception. The integration of animal neurobiology with human perception has led to important insights about how information-bearing parameters of sounds may be represented by combination-sensitive neurons in the brain (20, 21). However, it is still difficult to establish common theories about processing and perception of acoustic objects by the brain, because too few behavioral studies in mammals have explicitly identified information-bearing parameters of sounds, especially of communication sounds, for acoustical object perception (22–27).

Here, we present the behavioral (psychoacoustical) evidence about time-critical spectral integration for the perception of a communication call in a mammal. We show a close correspondence of acoustical object perception in mice and men, which suggests that studying mammalian communication and the related neural processes in animal models is likely to provide adequate descriptions of neural representation of complex sounds up to the perceptual level in the human brain.

Our study deals with the perception of mouse pup wriggling calls (Fig. 1*a*) by the pups' mothers in a natural communication situation. Wriggling calls are produced by pups struggling for the mother's nipples when she is in a nursing position on her litter (28). The calls reliably release pup-caring behavior in the mother (29). Wriggling calls have an average duration of 120 ms, cover a frequency range between about 2 and 20 kHz, and often consist of a main frequency component (fundamental frequency) near 4 kHz and a minimum of two overtones (Fig. 1*a*). This basic harmonic structure may be superimposed by rapid frequency shifts, amplitude modulations, and noises (24, 28, 29). The presence of three frequencies at 3.8, 7.6, and 11.4 kHz for 100 ms (Fig. 1*b*) is necessary and sufficient for wriggling call perception, i.e., for the release of maternal behavior (24). To stress the perceptual significance of these three frequencies and the correspondence of wriggling calls with human speech vowels, we will call them "formants" in this article. Formants define frequency contours of increased intensity (resonance frequencies of the vocal tract) important for vowel perception and discrimination in human speech. Here, we vary the total duration of synthesized three-formant call models and the onset and offset times of the first formant (fundamental frequency of the harmonic complex) relative to the higher ones. Thus, we test the acoustic boundaries in the time domain for grouping the three formants to a biologically meaningful acoustic object that leads, if heard in a sequence, to the release of maternal behavior.

Materials and Methods

Animals. Fifty-eight primiparous lactating house mice (*Mus domesticus*, outbred strain NMRI) aged 12–16 weeks, with their 1- to 5-day-old pups (litters standardized to 14 pups) were housed at an average 22°C and a 12-h light-dark cycle (light on at 7 h). They lived in plastic cages (26.5 × 20 × 14 cm) that had a circular hole (Ø 9 cm) covered by a fine gauze in the center of bottom of the cage. Wood shavings served as nest material. Food and water were available *ad libitum*.

Synthesis and Playback of Call Models. Results from our study about the perception of wriggling calls in the frequency domain (24) are the basis for the design of the frequency and time structure of the call models used here. All 17 wriggling call models were synthesized (PC 386, 125 kHz DA/AD card Engineering Design, SIGNAL SOFTWARE 2.2) to have the three frequencies: 3.8, 7.6, and 11.4 kHz (Fig. 1*b*). These frequencies were initiated at zero phase and had rise and fall times of 5 ms within the total

This paper was submitted directly (Track II) to the PNAS office.

*To whom reprint requests should be addressed. E-mail: guenter.ehret@biologie.uni-ulm.de.

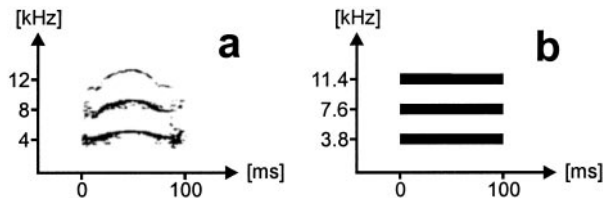


Fig. 1. (a) Spectrogram of a typical wriggling call of mouse pups (modified from ref. 28). (b) Synthesized standard call model (model A) with frequencies at 3.8, 7.6, and 11.4 kHz as a nearly optimum releaser of maternal behavior (24).

durations given below. In a first series of tests, call models were presented in which the two higher harmonics always had a duration of 100 ms and started and ended simultaneously (Fig. 2). The start of the fundamental frequency (3.8 kHz) was varied systematically with regard to the start of the two higher harmonics as shown in Fig. 2. It started simultaneously with the other formants (A), started +10, +20, +30, +35, +40, +50 ms earlier (B, C, D, E, F, G), or started -10, -20, -30, -40, -50 ms later (H, I, J, K, L). In all these cases, the fundamental frequency ended simultaneously with the two higher harmonics. In a second series of tests, the three frequencies in the call models were always of equal duration (Fig. 3). Although starting and ending simultaneously, they had durations of either 50, 70, or 80 ms (M, N, O). In call models P and Q (Fig. 3), all frequencies had a 100-ms duration, whereas the fundamental frequency either started 30 ms earlier (P) or 30 ms later (Q) than the second and third harmonics.

All call models were played back in bouts of four calls [bouts of two to five wriggling calls are most frequently produced by 1- to 5-day-old pups (24)] with intercall intervals of 200 ms relative to the start and end of the two higher harmonics. The call models were bandpass-filtered (3–12 kHz, Kemo VBF 8, 96 dB/octave,

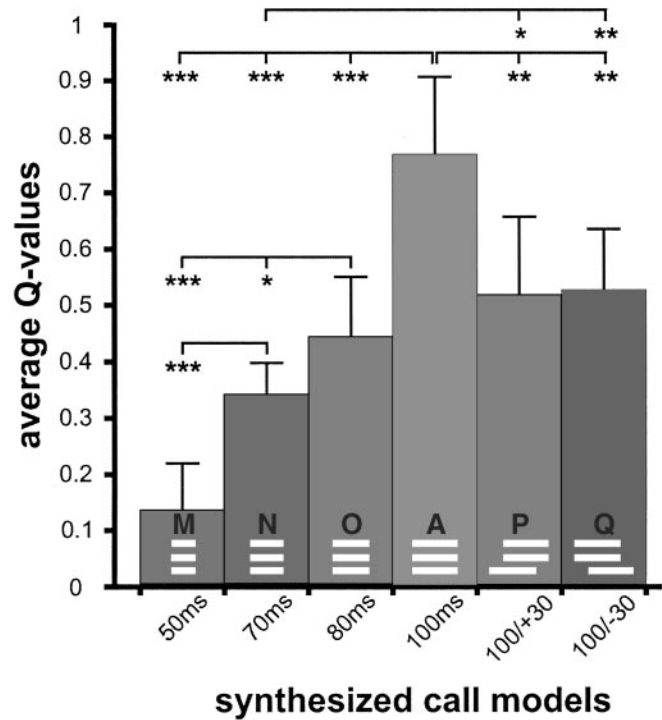


Fig. 3. Average Q value of the standard call model (A) compared with Q values of call models having shorter durations (M, N, O) or asynchronous on- and offsets of the fundamental frequency (P, Q). The structures of the call models are shown in, and the parameter values shown below the respective bars. Statistically significant differences between Q values are indicated with *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

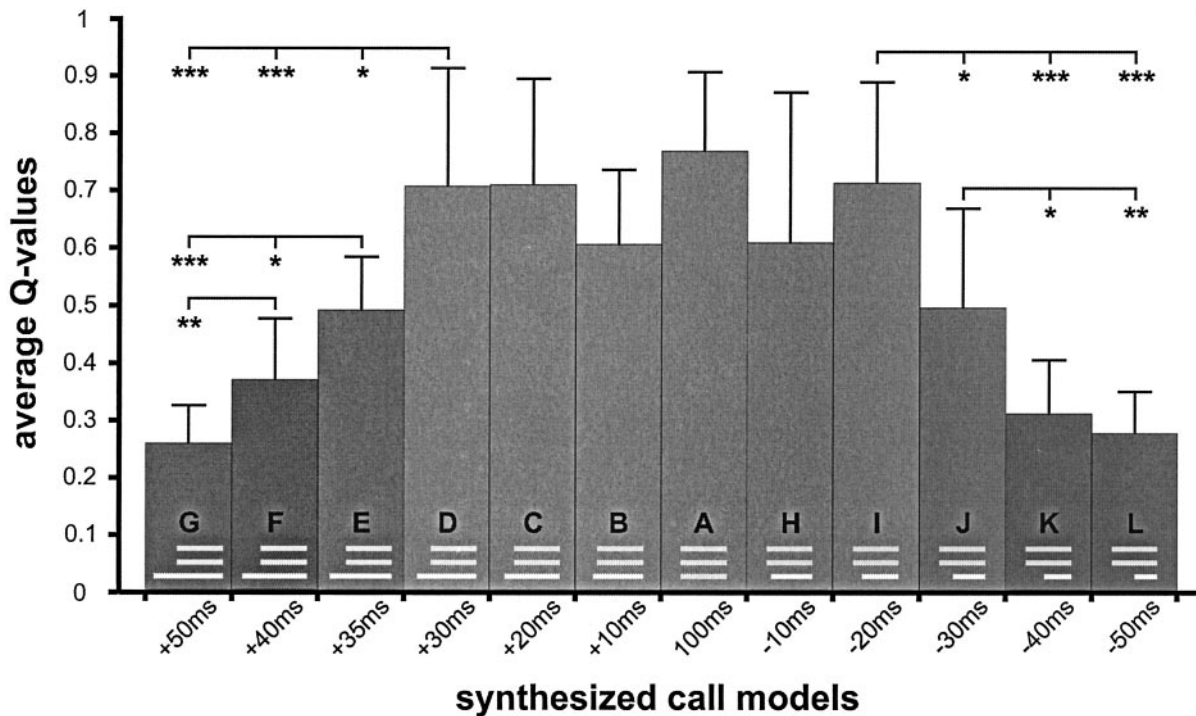


Fig. 2. Average Q values with standard deviations indicating the effectiveness of the synthesized call models (A–L) for the release of maternal behavior. The frequency and time structure of each call model is shown in the respective bars. The onset lead and lag times of the fundamental frequency are indicated below the bars. Statistically significant differences between Q values are indicated with *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

Ziegler Instruments, Mönchengladbach, Germany), amplified (Denon PMA-1060, Fröschl Elektro, Ulm, Germany), and sent to a loud speaker (Dynaudio D28, Dynaudio, Bensenville, IL) mounted independently of the cage 1 cm below the hole in the cage of the test mouse. The speaker had a flat ± 6 -dB frequency spectrum (Nicolet 446B spectrum analyzer) between 3 and 19 kHz measured in the cage. The total intensity of all call models was 70 dB sound pressure level in the nest area of the cage (Brüel & Kjaer microphone 4133 plus measuring amplifier 2636; Brüel & Kjaer Instruments, Marlborough, MA), which is close to the sound pressure level of natural wriggling calls reaching the ear of a nursing mother (28). Each frequency had the same sound pressure level (60.5 dB), all three adding up to 70 dB sound pressure level.

Recording and Analysis of Maternal Behavior. Observations were made under dim red light in a sound-proof and anechoic room between 9–12 and 17–20 h. The cage with the mother and her litter was suspended in the room at least 1 h before the observation started. The cover grid of the cage was removed and its height increased by a 6-cm-high plastic head piece. A video camera with microphone was fixed above the cage to observe and record outside the room the behavior of the mother and the sounds from the cage (natural wriggling calls of the pups, playbacks of call models).

Playbacks of one of the 17 call models were started only when the mother was in a nursing position on her litter and while the pups were not vocalizing wriggling calls. About 50 bouts of one call model were presented at intervals of 20–120 s, adding up to a 45-min total observation period. Maternal responses to the natural calls or synthesized call models were noted if the mother responded within 3 s after onset of the respective calls with either “licking of pups,” “changing of nursing position,” or “nest building” (29). Lack of responses of the mother and the number of bouts of wriggling calls produced by the litter were also noted. When the litter produced wriggling calls just when a playback of a call model had been started, the mother’s response was not considered. Also, if the mother responded with several maternal activities to natural or synthesized calls, only the first response was considered.

The effectiveness of the synthesized calls relative to the natural calls for the release of maternal behavior was expressed by a quality coefficient, Q . As explained (24), Q was calculated as the relative number of responses to the bouts of a given call model presented divided by the relative number of responses to bouts of the natural calls of the pups produced during the 45-min observation period. This calibration of the responsiveness to the call models eliminates influences of variations of actual motivation, attention, and arousal among the mothers during the tests and, thus, leads to comparable values (Q) of the effectiveness of different calls models for the release of maternal behaviors (24). If $Q = 0$, a mother did not respond to a given call model. If $Q = 1$, a mother responded with the same relative rate to a given call model as to the wriggling calls of her pups.

Ten mothers having 1-, 2-, 3-, 4-, or 5-day-old pups, respectively, were always tested with a given call model, and individual Q values were calculated. A mother was tested with a maximum of three different call models on different days. Despite this repeated testing of several animals, the results showed (see *Appendix A*) that the data obtained can be regarded as completely independent of each other. Hence, the responsiveness to different call models was compared statistically with tests that assume independence of the data [H-test analysis of variance, and U test, both two-tailed (30)].

Results and Discussion

Time Window for Auditory Object Perception. The effectiveness, expressed by the quality coefficient Q , of the 17 wriggling call

models to release maternal behavior is shown in Figs. 2 and 3. The standard is the call model A (all three formants in synchrony) reaching an average Q value of 0.76, which is very similar to the Q value of the same stimulus obtained in our earlier study (24) and does not differ significantly from Q values of natural calls played back from tape (24). Q values significantly smaller than that of call model A can be obtained by changing the composition or reducing the number of frequency components in the call models, or shifting the frequencies to a higher range (24). The importance of time windows and temporal coherence for spectral integration is shown here. An advancing presentation of the fundamental frequency (Fig. 2) up to +30 ms (call models B, C, and D) and a delaying presentation of up to –20 ms (call models H and I) have no significant influence on call perception (H test, $P > 0.1$). Lead times of the first formant of more than +30 ms (call models E, F, and G) and lag times of more than –20 ms (call models J, K, and L) decreased the responsiveness of the mothers continuously and significantly (H test, $P < 0.001$ in each case), compared with call models A–D or A, H, I, respectively (Fig. 2). The continuous decrease is emphasized by the significant differences between respective pairs of average Q values (U test, at least $P < 0.05$) as shown in Fig. 2.

Our data indicate grouping of the three formant frequencies (3.8, 7.6, and 11.4 kHz) to an optimum wriggling-call percept within a time window of +30-ms lead time and –20-ms lag time of the first formant relative to the others. Comparable results on time-critical perceptual groupings of frequency components to auditory percepts that differ significantly from other percepts have been shown in various studies on humans. For example, (i) a maximum onset lead time of about 32 ms occurs for formant integration in speech vowels (31, 32); (ii) a maximum onset lead or lag time of about 20 ms occurs for formant integration in a nonspeech context (33); (iii) a fusion of two frequencies to a single percept occurs only if both start within about 29 ms of each other (34); (iv) a critical time window of less than about 35-ms lead or lag time exists within which a frequency can be integrated as the second formant into the stream of a three-formant complex (35); (v) a minimum gap width of about 30 ms exists to produce normal sounding consonants in a vowel sequence of speech (36); (vi) a maximum onset lead time of about 40 ms of a tone occurs to produce a maximum pitch shift in a later starting frequency complex (37); and (vii) a minimum gap width of almost 40 ms occurs between a tone ending before a frequency complex and the complex to prevent a pitch shift of the complex (38).

These studies just mentioned and further studies on the detection of gaps in a sequence of different frequencies (39), the effect of gaps for the perception of auditory streams (40), and the discrimination of phonemes by voice-onset time in speech both by humans (41, 42) and animals such as budgerigars (43), chinchillas (25, 26), and rhesus monkeys (27) all show that spectral integration across frequency channels may lead to a particular auditory percept only if all frequency components to be grouped together start within a ± 20 - to 30-ms time window or are separated from other frequency components by 20–40 ms. Thus, a common critical time window exists in mice, other mammals, birds, and humans for spectral integration or fusion of frequency components within the time window (formation of an acoustic object in the frequency domain) and for the separation of frequency components outside this window (separation of acoustic objects).

Auditory Scene Analysis and Streaming. If mouse mothers perceive only a single wriggling call of a favorable frequency composition and time structure for the release of maternal behavior, they do not respond. Maternal behavior is elicited only if the wriggling calls (the acoustic objects) are heard in a sequence of at least two

calls (G.E. and S. Riecke, unpublished results). Hence, the condition for the perception of the biological meaning of the wriggling calls is a stream of adequate acoustical objects, i.e., an acoustical Gestalt. In our tests we offered such a stream by always presenting four call models of the same type in a sequence. In terms of auditory scene analysis (4), wriggling calls must be perceived as a single stream of adequate acoustical objects, such as call models A–D, H, and I, to reach biological significance. In call models E–G and J–L, our mice seem to have perceived the streams of two acoustic objects, one stream representing the sequence of the two higher harmonics, and the other stream representing the sequence of the fundamental frequency of the call models. Both streams are not adequate to release maternal behavior; both have an insufficient acoustical Gestalt. The Q values obtained (Fig. 2) are nearly equal to the sum of the Q values from responses to the two higher harmonics (presented together) and to a single frequency as shown in the previous study (24). These indications of auditory scene analysis and the perception of auditory streams in mice add significantly to the few behavioral studies on auditory scene analysis in animals available for goldfish (44, 45) and starlings (46, 47), because those animals were trained to discriminate between auditory patterns representing one or two auditory streams, whereas our mice were not trained at all and discriminated acoustical patterns in a natural communication situation.

Perception of Call Duration. Because natural wriggling calls consist of a minimum of three harmonically related formants (Fig. 1*a*) of a duration of 120 ± 40 ms (mean \pm standard deviation) (24, 28), it is conceivable that call models A–D, H, and I with a synchronous duration of the three formants of at least 80 ms are the most effective releasers of maternal behavior [Q values are not significantly different from the Q values obtained with natural calls played back from a tape (24)]. In the studies cited (24, 31–35, 42), the critical durations of time windows for spectral integration depended on various other parameters such as the experience of listeners for the task, the presentation of the stimuli under high- or low-uncertainty conditions, the spacing of the frequencies to be integrated (e.g., whether frequency components can be spectrally resolved or not), and the total durations of the sound elements. In this study, we controlled these parameters, i.e., the mice were not trained for the task, our call models were presented under high-uncertainty conditions, and the formant frequencies could be resolved by the critical band mechanism of the mouse ear (48), and tested how the grouping of formants to an adequate wriggling-call percept depended on the duration of the simultaneous presence of the formants. The results are shown in Fig. 3. A shortening of the call duration from 100 ms down to 50 ms (call models A, O, M, and N) significantly decreased the effectiveness of the calls for the release of maternal behavior (H test, $P < 0.01$). The increase of Q with increasing call duration (d) is linear ($Q = 0.0126d - 0.5229$; correlation coefficient $r = 0.9152$; $P < 0.001$; 38 df) and the Q values of all duration steps differ from each other (Fig. 3; U test, $P < 0.05$ at least).

The duration in which all formants are present simultaneously interferes with the duration in which single formants are present, and with the synchrony of on- and offsets. This interference is evident from Q values of call models P and Q, which are significantly larger than the Q value of call model N (Fig. 3), similar to the Q value of call model J, and significantly smaller than the Q value (U test, $P < 0.05$) of call model D (Fig. 2). Combined effects of total duration and on- and offset asynchronies on the inclusion of frequency components in acoustic streams have been shown for humans (35). Our data indicate a nonlinear interaction in the perception of call duration and onset asynchronies by a lag of the first formant. Call models I, J, and L (Fig. 2) have significantly larger Q values than call models O,

N, and M, respectively (I vs. O, $P < 0.01$; J vs. N, $P < 0.05$; L vs. M, $P < 0.01$; U test, two-tailed) although the durations of the simultaneous presence of the three formants are the same, namely 80 ms (I, O), 70 ms (J, N), and 50 ms (L, M). Further tests will have to show which durations of simultaneous presence of all three formants are necessary for the fusion to an acoustic object and the formation of one auditory stream depending on the time lag of the first or the other formants. In the longer-onset delays of a formant (call models J–L; Fig. 2), results from humans (4, 35, 37) suggest that more than 80 ms of simultaneous presence, as in call model I (Fig. 2), is necessary for a perceptual fusion of the formants and streaming of an optimum wriggling call.

The continuous type of dependence of wriggling-call perception on call duration (Fig. 3, call models A, O, N, and M) is clearly different from categorical perception of ultrasonic calls by unconditioned mice in a natural communication situation, in which preferred calls are discriminated from nonpreferred ones at a sharp (only 5 ms wide) boundary of about 27-ms duration (49). The very different strategies of call discrimination and perception with regard to call duration for wriggling calls and ultrasound in mice stress the fact that auditory percepts in the time domain do not simply reflect peripheral processes and resolution but require central mechanisms of integration that select time intervals and durations of communicative importance (see discussion in refs. 4, 9, 11, 13, 14, 20, 21, 38, and 39). The neural selection processes may be determined mainly by the evolution of ecological adaptations (refs. 20, 21; present work) or by cognitive strategies evident in speech perception (50, 51).

Neural Correlates of Perception in the Time Domain. The auditory midbrain (inferior colliculus) most probably is the first level of the auditory pathway with neurons being selective for sound duration (52–54). Some of the neurons in the inferior colliculus show duration tuning that is adapted to the acoustical ecology of the species. Thus, neurons in bats prefer short-duration sounds of 1–30 ms that are compatible with their echolocation calls (53, 55–57), whereas neurons in frogs are sensitive to longer sounds compatible with mating calls (58). A related study of the mouse inferior colliculus shows that almost 70% of duration-sensitive neurons prefer long-duration sounds (59). In the auditory cortex, duration tuning is extended to durations of 100 ms in bats (60) and more than 200 ms in cats (61). Further, a considerable number of neurons have a long-pass characteristic with little responsiveness to durations shorter than about 100 ms and strong responses to longer-duration sounds (60, 61). If such long-pass neurons exist in the mouse auditory cortex, they could be the basis for the preference of wriggling-call models of 100-ms duration (call model A) against calls of shorter durations (call models M–O) in the present study.

A considerable body of literature on combination-sensitive neurons in the inferior colliculus, medial geniculate body, and auditory cortex has accumulated that cannot be reviewed here in detail. Neurons have been found to show various kinds of selectivity to combinations of frequency components or formants and to combinations of spectral and temporal patterns that all may be important information-bearing parameters for guiding sound-based behavior in animals (see refs. 6–13, 15–19, 62–68; reviews in refs. 14, 20, 21, and 52). For example, combinations of formants and formant transitions, together with a range of delays between pulse and echo, encode within various neural responses the relative target velocity and target distance, respectively (15–17, 62–66). Neurons in the auditory cortex of bats (13, 14) and higher auditory areas in the avian brain (7–10) are syntax-sensitive, i.e., they respond preferentially to the temporal order of spectrally defined elements of species-specific communication sounds or bird's own song. Such selective responses concerning the perceived acoustical Gestalt, i.e., the

auditory scene or auditory streams, are shaped by central mechanisms of neural facilitation and inhibition. In short, inhibitory influences between frequency components of a sound stream (e.g., observed as forward masking) seem to dominate over times of about 50 ms or less, whereas facilitatory influences (e.g., seen as forward enhancement) become important at times of about 50–200 ms (69–73). In view of these data on neural mechanisms, our present results on wriggling-call perception may be interpreted in the following way. The fusion of the formants of a wriggling call to an acoustic object was disturbed by inhibitory influences of one or two formants starting earlier than the other formant(s) of the call for call models E–G, J–L, P, and Q. Such inhibitory influences from onset disparities of the formants were not large enough to alter significantly the percepts of call models B–D, H, and I compared with call model A. The intervals of 200 ms each between the four calls in a sequence ensured facilitatory influences of one call to the next so that the call sequence could be perceived as a biologically significant stream of relevant acoustic objects.

Appendix A

We tested in a correlation analysis (30) whether the second Q value obtained from animals used in two tests depended on the

type of call model (relevant models such as call models A–D, H, and I, or irrelevant models such as all of the other models shown in Figs. 2 and 3) used in the first test of the same animals, and whether the third Q value obtained from animals used in three tests depended on the type of call models used in the two previous tests of the same animals. In none of the 11 cases did we find a statistically significant correlation (correlation coefficients r always indicated significance levels of $P > 0.3$). Hence, Q values of repeatedly tested animals are independent data. In addition, a post hoc analysis showed that if we had used only one Q value of every animal and, thus, worked with initially independent data, we had a minimum of six and a maximum of nine independent Q values in the groups of tested call models (Figs. 2 and 3). On the basis of such a smaller data sample, all statistical evaluations had the same results as indicated in the main text with significance levels of at least $P < 0.05$. Because we could establish the independence of all our tests (see above), we built our results and conclusions on all of the data obtained in this study and shown in Figs. 2 and 3.

We thank Dr. K.-H. Esser for valuable help in preparing the call models. This work was supported by the Deutsche Forschungsgemeinschaft (Eh 53/17-1,2).

- Hirsh, I. J. & Watson, T. S. (1996) *Annu. Rev. Psychol.* **47**, 461–484.
- Rasch, R. A. (1978) *Acoustica* **40**, 21–33.
- Darwin, C. J. (1981) *Q. J. Exp. Psychol.* **33A**, 185–207.
- Bregman, A. S. (1990) *Auditory Scene Analysis* (MIT Press, Cambridge, MA).
- Yost, W. A. & Sheft, S. (1993) in *Human Psychophysics*, eds. Fay, R. R. & Popper, A. N. (Springer, New York), pp. 193–236.
- Fuzessery, Z. M. & Feng, A. S. (1982) *J. Comp. Physiol. A* **146**, 471–484.
- Scheich, H., Langner, G. & Bonke, D. (1979) *J. Comp. Physiol. A* **132**, 257–276.
- Langner, G., Bonke, D. & Scheich, H. (1981) *Exp. Brain Res.* **43**, 11–24.
- Margoliash, D. & Fortune, E. S. (1992) *J. Neurosci.* **12**, 4309–4326.
- Doupe, A. J. (1997) *J. Neurosci.* **17**, 1147–1167.
- Wang, X., Merzenich, M. M., Beitel, R. & Schreiner, C. E. (1995) *J. Neurophysiol.* **74**, 2685–2706.
- Ohlemiller, K. K., Kanwal, J. S. & Suga, N. (1996) *NeuroReport* **7**, 1749–1755.
- Esser, K.-H., Condon, C. J., Suga, N. & Kanwal, J. S. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14019–14024.
- Kanwal, J. S. (1999) in *The Design of Animal Communication*, eds. Hauser, M. D. & Konishi, M. (MIT Press, Cambridge, MA), pp. 133–157.
- Suga, N., O'Neill, W. E. & Manabe, T. (1978) *Science* **200**, 778–781.
- Suga, N., O'Neill, W. E., Kuzirai, K. & Manabe, T. (1983) *J. Neurophysiol.* **49**, 1573–1626.
- Fitzpatrick, D. C., Kanwal, J. S., Butman, J. A. & Suga, N. (1993) *J. Neurosci.* **13**, 931–940.
- Paschal, W. G. & Wong, D. (1994) *J. Neurophysiol.* **72**, 366–379.
- Condon, C. J., Galazyuk, A., White, K. A. & Feng, A. S. (1997) *Aud. Neurosci.* **3**, 269–287.
- Suga, N. (1988) in *Auditory Function. Neurobiological Bases of Hearing*, eds. Edelman, G. M., Gall, W. E. & Cowan, W. M. (Wiley, New York), pp. 679–720.
- Suga, N. (1996) in *The Cognitive Neurosciences*, ed. Gazzaniga, M. S. (MIT Press, Cambridge, MA), pp. 295–313.
- Ehret, G. & Haack, B. (1982) *J. Comp. Physiol. A* **148**, 245–251.
- May, B., Moody, D. B. & Stebbins, W. C. (1988) *Anim. Behav.* **36**, 1432–1444.
- Ehret, G. & Riecke, S. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 479–482.
- Kuhl, P. K. & Miller, J. D. (1975) *Science* **190**, 69–72.
- Kuhl, P. K. & Miller, J. D. (1978) *J. Acoust. Soc. Am.* **63**, 905–917.
- Waters, R. S. & Wilson A. (1976) *Percept. Psychophys.* **19**, 285–289.
- Ehret, G. (1975) *Behavior* **52**, 38–56.
- Ehret, G. & Bernecker, C. (1986) *Anim. Behav.* **34**, 821–830.
- Sachs, L. (1999) *Angewandte Statistik* (Springer, Berlin).
- Darwin, C. J. (1984) *J. Acoust. Soc. Am.* **76**, 1636–1647.
- Darwin, C. J. & Sutherland, N. S. (1984) *Q. J. Exp. Psychol.* **36A**, 193–208.
- Pisoni, D. W. (1977) *J. Acoust. Soc. Am.* **61**, 1352–1361.
- Bregman, A. S. & Pinker, S. (1978) *Can. J. Psychol.* **32**, 19–31.
- Dannenbring, G. L. & Bregman A. S. (1978) *Percept. Psychophys.* **24**, 369–376.
- Huggins, A. W. F. (1972) *J. Acoust. Soc. Am.* **51**, 1270–1278.
- Darwin, C. J. & Ciocca, V. (1992) *J. Acoust. Soc. Am.* **91**, 3381–3390.
- Ciocca, V. & Darwin, C. J. (1999) *J. Acoust. Soc. Am.* **105**, 2421–2430.
- Phillips, D. P., Tailor, T. L., Hall, S. E., Carr, M. M. & Mossop, J. E. (1997) *J. Acoust. Soc. Am.* **101**, 3694–3705.
- Dannenbring, G. L. & Bregman, A. S. (1976) *J. Acoust. Soc. Am.* **59**, 987–989.
- Lieberman, A. M., Paris, K. S., Kinney, J. A. & Lane, H. (1961) *J. Exp. Psychol.* **61**, 379–388.
- Summerfield, Q. & Haggard, M. (1977) *J. Acoust. Soc. Am.* **62**, 435–448.
- Dooling, R. J. (1989) in *Comparative Psychology of Audition: Perceiving Complex Sounds*, eds. Dooling, R. J. & Hulse, S. H. (Erlbaum, Hillsdale, NJ), pp. 423–444.
- Fay, R. R. (1998) *Hear. Res.* **120**, 69–76.
- Fay, R. R. (2000) *J. Assoc. Res. Otolaryngol.* **1**, 120–128.
- Hulse, S. H., MacDougall-Shackleton, S. A. & Wisniewski, A. B. (1997) *J. Comp. Psychol.* **111**, 3–13.
- MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q. & White, W. (1998) *J. Acoust. Soc. Am.* **103**, 3581–3587.
- Ehret, G. (1976) *Biol. Cybernet.* **24**, 35–42.
- Ehret, G. (1992) *Anim. Behav.* **43**, 409–416.
- Merzenich, M. M., Schreiner, C., Jenkins, W. & Wang, X. (1993) *Ann. N.Y. Acad. Sci.* **682**, 1–22.
- Kuhl, P. K. (1999) in *The Design of Animal Communication*, eds. Hauser, M. D. & Konishi, M. (MIT Press, Cambridge, MA), pp. 419–450.
- Feng, A. S., Hall, J. C. & Goller, D. M. (1990) *Prog. Neurobiol.* **34**, 313–329.
- Casseday, J. H., Ehrlich, D. & Covey, E. (1994) *Science* **264**, 847–850.
- Covey, E. & Casseday, J. H. (1999) *Annu. Rev. Physiol.* **61**, 457–476.
- Ehrlich, D., Casseday, J. H. & Covey, E. (1997) *J. Neurophysiol.* **77**, 2360–2372.
- Jen, P. H. S. & Feng, R. B. (1999) *J. Comp. Physiol. A* **184**, 185–194.
- Jen, P. H. S. & Zhou, X. M. (1999) *J. Comp. Physiol. A* **185**, 471–478.
- Gooler, D. M. & Feng, A. S. (1992) *J. Neurophysiol.* **67**, 1–22.
- Brand, A., Urban, A. & Grothe, B. (2000) *J. Neurophysiol.* **84**, 1790–1799.
- Galazyuk, A. V. & Feng, A. S. (1997) *J. Comp. Physiol. A* **180**, 301–311.
- He, J., Hashikawa, T., Ojima, H. & Kinouchi, Y. (1997) *J. Neurosci.* **17**, 2615–2625.
- Olsen, J. F. & Suga, N. (1991) *J. Neurophysiol.* **65**, 1275–1296.
- Mittmann, D. H. & Wenstrup, J. J. (1995) *Hear. Res.* **90**, 185–191.
- Wenstrup, J. J. (1999) *J. Neurophysiol.* **82**, 2528–2544.
- Portfors, C. V. & Wenstrup, J. J. (1999) *J. Neurophysiol.* **82**, 1326–1338.
- Misawa, H. & Suga, N. (2001) *Hear. Res.* **151**, 15–29.
- Yan, J. & Suga, N. (1996) *Hear. Res.* **93**, 102–110.
- Sutter, M. L. & Schreiner, C. E. (1991) *J. Neurophysiol.* **65**, 1207–1226.
- Calford, M. B. & Semple, M. N. (1995) *J. Neurophysiol.* **73**, 1876–1891.
- Brosch, M. & Schreiner, C. E. (1997) *J. Neurophysiol.* **77**, 923–943.
- Brosch, M., Schulz, A. & Scheich, H. (1999) *J. Neurophysiol.* **82**, 1542–1559.
- Brosch, M. & Schreiner, C. E. (2000) *Cereb. Cortex* **10**, 1155–1167.
- Fishman, Y. I., Reser, D. H., Arezzo, J. C. & Steinschneider, M. (2001) *Hear. Res.* **151**, 167–187.