Supporting Text

Supporting Analysis: The role of adaptation

In the main experiment, if adaptation occurs and both target and tones are effectively at a lower level, then there should still be trading between tones and vowel; it should just warp the trading between the effective attenuation in the tones and vowel. This is not what is observed.

Specifically, under the assumption that there is significant adaptation, the "effective attenuation of the target in the vowel" should be relative to the unadapted target level (as the comparison is of the spectral shape, and presumably the vowel components, with 240 ms of silence between them, show little or no adaptation). In contrast, the "effective attenuation of the target in the tones" should be relative to the target level after adaptation (as the comparison is to the other tones, which will show adaptation as well).

Then, if there is adaptation, and the trading hypothesis holds:

$$1 = 10^{\frac{-Att_{tones}}{10}} + 10^{\frac{Att_{adaptation}}{10}} 10^{\frac{-Att_{vowel}}{10}}$$

where

$Att_{adaptation}$ is the effective attenuation of the target (in dB) due to adaptation

Att_{tones} is the target attenuation in the tones (in dB) due to trading with vowel

Att_{tones} is the target attenuation in the vowel (in dB) due to trading with tones

This leads to a different form of trading relationship, which depends on the effective amount of the target attenuation due to adaptation, in dB. Figure 3 shows these trading contours for different assumed levels of adaptation (in 3 dB steps).

The figure shows that the "orphan" does not fall on the same contour as the other data. While the three other points fall between the "adaptation trade" contours for 3 and 6 dB of adaptation, the orphan falls on the 9 dB adaptation curve.

Thus, while adaptation may contribute to the "energy loss" we find, it cannot explain the orphan case. Quantitatively, while adaptation may contribute to the observed results, it cannot account fully for the perceptual loss of the *target* when the spatial cues of the *target* match those of the *vowel* and differ from those of the *tones.*

Supplemental Experiment

As in the main experiment, stimuli in the supplemental experiment consist of two objects (here, the *simultaneous complex* and the *complex stream*, taking on the roles of the *vowel* and *tones* of the main experiment, respectively) that compete for ownership of an ambiguous *target*. In this supplemental experiment, listeners matched the perceived spectral makeup of the *simultaneous complex* and the *complex stream* (see Supporting Methods). As in the main experiment, the *simultaneous complex* is always presented from in front of the listener, while the *target* and the *complex stream* either came from in front or to the side of the listener.

Figure 4B shows the mean attenuations of the *target* (averaged across subjects) that produce perceptual matches to the *simultaneous complex* and *complex stream* spectral content, plotted against each other in a format comparable to that of Fig. 3. As in the main experiment, changes in the spatial configuration of the sound elements alter the perceived spectral content of the objects. When the spatial cues of the *target* and *complex stream* match and the *target* and *simultaneous complex* do not match (Fig. 4B, filled triangle), the *target* contributes a great deal to the *complex stream*, but contributes little to the *simultaneous complex*. When the *target* location matches both the location of the *simultaneous complex* and of the *complex stream*, the *target* contributes strongly to the *simultaneous complex* and weakly to the *complex stream* (Fig. 4B, filled circle). When the *target* location matches neither *simultaneous complex* nor *complex stream* locations, its contribution to the *simultaneous complex* increases and its contribution to the *complex stream* decreases (Fig. 4B, open triangle). Finally, when the *target* location matches that of the *simultaneous stream* and does not match that of the *complex stream*, it contributes almost nothing to the perceived content of the *complex stream,* but contributes significantly to the perceived content of the *simultaneous stream* (Fig. 4B, open circle)*.*

These results differ from those of the main experiment in a number of ways. First, in this experiment, results closely follow predictions based on an energy-trading hypothesis (solid curve in Fig. 4). Second, the *target* contributes strongly to the perceived content of the *simultaneous complex* in this experiment, unlike in the main experiment. Finally, there is no “lost target element” in this supplemental experiment: when the *target* location matches the location of the *simultaneous complex* and does not match the location of the *complex stream*, the *target* is heard as part of the *simultaneous complex* and does not contribute to the perceived content of the *complex stream.*

The repetition rates of the stimuli in the supplemental experiment were the same as in the main experiment. If the *tones* of the main experiment cause adaptation that reduces the internal level of the *target*, the *complex stream* in this supplemental experiment should cause similar adaptation of the *target* and decrease its perceptual contribution to the *simultaneous complex.* Instead, the *target* contributes strongly to the *simultaneous complex* in most conditions. Therefore, adaptation cannot account for the “lost *target*” in the main experiment. Similarly, when the *target* and *complex stream* have different locations, spatial cues change as rapidly here as in the main experiment. While it is possible that these rapid changes produce a spatially diffuse *target*, the *target* contributes strongly to the perceived content of the *simultaneous complex* of this supplemental experiment. Thus, binaural sluggishness cannot account for the “lost *target*” in the main experiment.

We attribute the difference in the pattern of results between the supplemental experiment and the main one to the fact that the *target* consists of multiple harmonically related components. Because both the *target* and *complex stream* are multi-tone harmonic complexes, rather than a single-frequency tone (see Fig. 4A), the balance between sequential and simultaneous grouping cues shifts compared to in the main experiment. This leads to stronger simultaneous grouping cues here than in the main experiment, and results in the elimination of the non-allocation effect.

Supporting Methods

Stimuli

Stimuli consisted of a 3-s long sequence, composed of ten identical repetitions of three 100-ms-long elements: two harmonic complexes of fundamental frequency 300 Hz (*complex stream*) followed by a *simultaneous complex* with fundamental frequency 200 Hz (see Fig. 4A). The *target* was a harmonic complex identical to those making up the *complex stream*, but presented simultaneously with the *simultaneous complex*. The *target* and *complex stream* bursts consisted of harmonics at 300, 600, 900, 1200, and 1500 Hz. The *simultaneous complex* consisted of harmonics 200, 400, 800, 1000, 1400, and 1600 Hz. The *complex stream, target*, and *simultaneous complex* all were gated with a Blackman window (60-ms duration) and were separated in time by a silent gap of 40 ms. The sequence of repeating *complex stream* and *simultaneous complex* caused a percept of two distinct auditory objects (a rapidly repeating complex with a pitch of 300 Hz and a complex repeating at one third that rate).

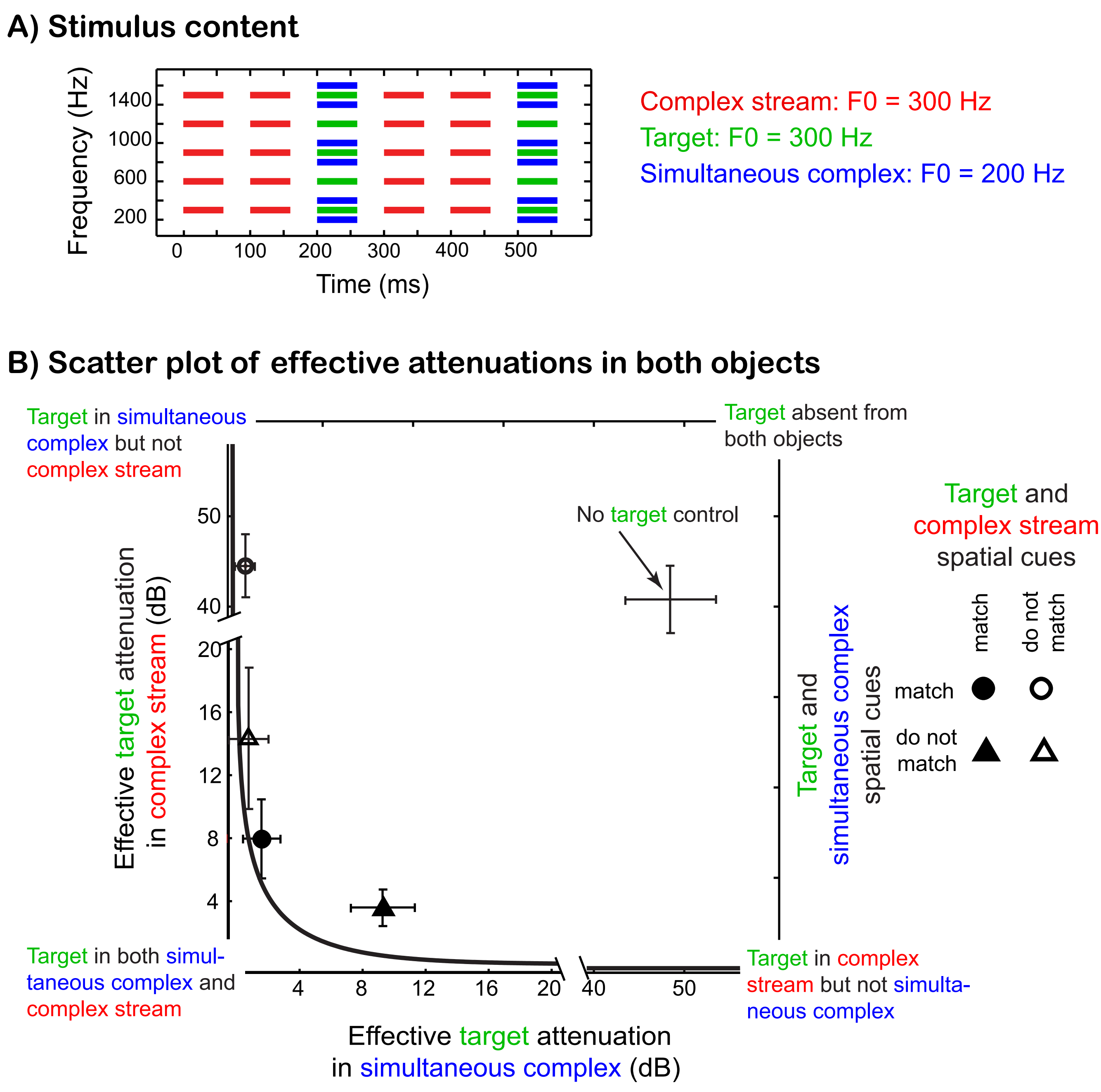
This stimulus design caused a distinct change in both the spectral density and the perceived pitch of the *simultaneous complex* depending on whether or not the *target* was heard as part of the *simultaneous complex*. When the *target* was heard in the *simultaneous complex*, listeners perceived a complex with a dense spectral composition and a pitch of 100 Hz (corresponding to the missing fundamental). When the *target* was not heard as part of the *simultaneous complex,* its perceived pitch was an octave higher (200 Hz) and it had a more sparse spectral density.

Spatial cues for the *target*, *simultaneous complex*, and *complex stream* were generated using the same head-related transfer functions as in the main experiment.

Procedures

As in the main experiment, two-object stimuli were presented in two blocks of trials differing only in the instructions to the subjects. In *complex stream* blocks, subjects matched the perceived content of the *complex stream*. In *simultaneous complex* blocks, subjects matched the perceived content of the *simultaneous complex*. Trials for each condition were presented in a different random order for each block of trials.

In both kinds of blocks, listeners used the method of adjustment to match the perceived content of the attended object. Each trial began by presenting a three-second-long test stimulus. This was followed by a three-second-long, single-object matching stimulus that consisted of an adjustable-level *target* and either a fixed-level *complex stream* or a fixed-level *simultaneous complex.* During presentation of the matching stimulus, subjects could adjust (in real time) the attenuation of the *target* by pressing one button to increase the *target* attenuation and a different button to decrease the *target* attenuation. Three-second-long test and matching stimuli alternated until the subject was satisfied with the perceptual match between the perceived content of the attended object in the two-object test stimulus and the content of that object in the single-object matching stimulus. When the subject was satisfied with the match, she/he pressed a third button, which stored the results of that trial and initiated the next trial in the block. Ten subjects participated in the supplemental experiment.


A) Stimulus content
Frequency (Hz)
1400
1000
600
200
0
100
200
300
400
500
Time (ms)
Complex stream: F0 = 300 Hz
Target: F0 = 300 Hz
Simultaneous complex: F0 = 200 Hz
B) Scatter plot of effective attenuations in both objects
Target in simultaneous complex but not complex stream
Target absent from both objects
Target in both simultaneous complex and complex stream
Target in complex stream but not simultaneous complex
No target control
Effective target attenuation in complex stream (dB)
50
40
20
16
12
8
4
4
8
12
16
20
40
50
Effective target attenuation in simultaneous complex (dB)
Target and simultaneous complex spatial cues
Target and complex stream spatial cues
match
do not match
match
do not match