# The scale illusion detection task: Objective assessment of binaural fusion in normal-hearing listeners

NICLAS A. JANßEN[1,*], LARS BRAMSLØW[2], SØREN RIIS[3], AND JEREMY MAROZEAU[1]

[1] *Hearing Systems, Department of Electrical Engineering, Technical University of Denmark, Kgs. Lyngby, Denmark*

[2] *Eriksholm Research Centre, Snekkersten, Denmark*

[3] *Oticon Medical, Smørum, Denmark*

The normal auditory system can fuse sounds from both ears into a single sound object (binaural fusion). This ability can be assessed subjectively by asking whether listeners perceive one or two sounds or by the scale illusion percept described by Deutsch (1975). The aim of the current study is to develop an objective task to measure binaural fusion. Twelve normal-hearing participants had to detect one deviant note within a stream composed of a repeating melody while simultaneously being presented with another stream of randomized notes. The experiment included 3 conditions. First, in a monaural condition both streams were presented to the same ear. Then, in a binaural condition every second note from each of the two streams was presented to the other ear. Finally, in a binaural control condition, the timbre of all the notes presented to one ear was altered severely, to prevent binaural fusion. The expected result was a better detection of deviant notes for listeners that are able to fuse streams across the two ears. Each condition had 24 repetitions. In the binaural and monaural conditions, average performance was about 80% correct, while the control condition showed a significantly lower performance of about 50%. Thus, this type of experiment can be used to test objectively if fusion takes place. It lays the foundation for further studies with bilateral and bimodal cochlear implant listeners.

## INTRODUCTION

While listening with both ears, humans fuse sounds binaurally into a common percept. Consequently, the voice of one speaker is perceived as one sound object, rather than two separate voices and information from both ears can be utilized to localize sounds or achieve superior speech perception in noise (cf. Middlebrooks *et al.*, 2017).

This ability to fuse binaurally presented sounds into one percept has been demonstrated elegantly by the scale illusion percept described by Deutsch (1975). It is based on a complex stimulation pattern, which consists of two melodies at two different frequency ranges (one high and one low). Both melodies go up and down in frequency over eight notes (see Fig. 1). These are presented in such a way that every

*Corresponding author: najan@elektro.dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 6: Adaptive Processes in Hearing, August 2017, Nyborg, Denmark. Edited by S. Santurette, T. Dau, J. C.-Dalsgaard, L. Tranebjærg, T. Andersen, and T. Poulsen. The Danavox Jubilee Foundation, 2017.
 ISBN: 978-87-990013-6-1.

second note from each stream is played to the other ear. Yet, listeners most often perceive the two ordered melodies, each of them lateralized to one side.

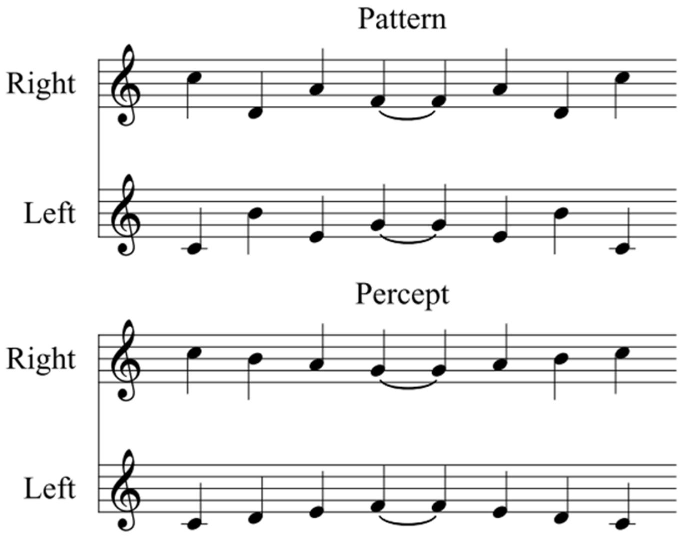


**Fig. 1:** Melodic pattern and percept for the original scale illusion (Deutsch, 1975).

These results clearly illustrate the ability to fuse sounds binaurally, as there is no other explanation for the reported percepts. However, the scale illusion experiment was based on a subjective task. Participants had to describe and draw their percepts. Such descriptions are open to interpretation by the experimenter and carry the risk of a potential biases. The aim of the current study is to find a fast and objective measure for binaural fusion by using a forced-choice detection paradigm.

## PARTICIPANTS

The new experiment has been evaluated with 12 normal-hearing listeners. They have been recruited from the students and staff at the Technical University of Denmark. Their age ranges from 24 to 31 years with a mean of about 26.95 and a standard deviation of about 1.73 years. Of the participants, 25.0% were female, 58.3% reported musical experience such as regular singing or playing an instrument and 75.0% were right-handed.

All participants provided informed consent and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

## STIMULI

The experiment was based on the scale illusion percept, featuring notes from the same set of frequencies. Unless denoted otherwise, these notes were presented as pure tones using the corresponding frequencies from the western tuning of the twelve-tone equal temperament scale. The duration of the notes was 250 ms, as in the original study. Further, half Hann-window ramps of 10 ms were applied to onset and offset of the notes to prevent spectral splatter. All the stimuli were loudness balanced using an adjustment procedure. This balancing has been performed twice for all frequencies used in the experiment, based on an external reference sound with preferred most comfortable loudness level chosen by the participant.

Stimuli were presented in a double-walled sound isolated listening booth via Sennheiser HDA-200 headphones.

## METHOD

Like in the original scale illusion experiment by Deutsch (1975), our new experiment used two melodies. We will refer to these two as the target stream and the distractor stream (or melody). The target stream was meant to be followed by the listener. It was chosen to be identical to one of the eight-note melodies in the scale illusion, either of the higher or lower frequency range. The distractor stream then consists of eight notes with frequencies chosen randomly from the other frequency range (i.e., if the target stream was of the higher frequency range, the distractor was of the lower range and vice versa). Both streams together were presented in the same way as in the original illusion, i.e., every second note from each stream is presented to the other ear in a pattern symmetrical to the middle of the eight notes sequence (see Fig. 2).

If a listener could fuse, he or she would be able to follow the target melody stream described above, while perceiving the distractor stream on the other side. If the listener was not able to fuse, he or she would perceive random melodies on both sides.

The task of the experiment was to detect a deviant note introduced into the target stream. A listener who fuses the binaural input into one stream was expected to be able to detect this deviant easily, whereas somebody who does not fuse the binaural input will fail to detect the change due to the random input in both ears.

First in each trial, the participant was presented with the target stream alone twice, to signal which stream to listen for. After that, target and distractor streams were presented six times. Of these six repetitions, the first three serve the purpose to allow for a build-up of streaming, as it has been reported that streamed percepts arise gradually over several seconds (Bregman, 1994, cf. Fig. 3). One of the last three repetitions contains the deviant note, as depicted exemplarily in interval B of Fig. 2. The currently presented interval was indicated on the graphical user interface. This arrangement thus represented a 3-alternative forced-choice paradigm. This prevents a bias in answers, compared to a yes/no task, since participants know that the deviant has occurred in one of the intervals (Wickens, 2002).

The deviant could occur at any of the inner 6 notes within a repetition of the melody, but not the first or last position, to prevent confusion of the interval. Its occurrence was counterbalanced with respect to the side of presentation, interval, higher or lower melody as the target stream and the order of trials has been randomized per participant to avoid order effects.

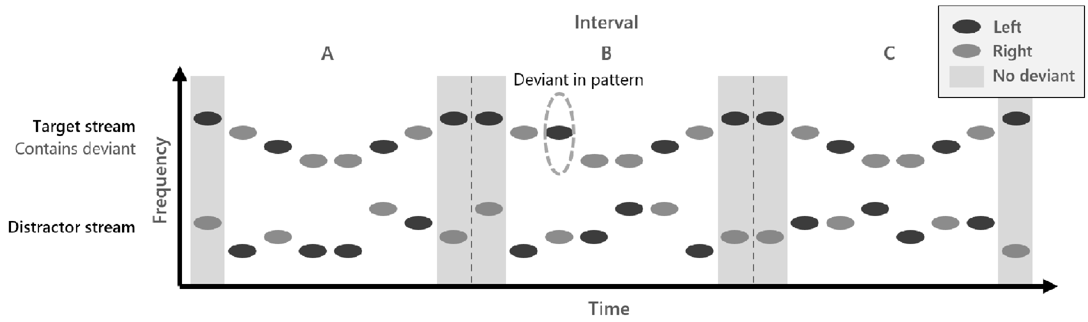


**Fig. 2:** Pattern of the scale illusion detection task, based on Deutsch (1975). A random deviant occurs in the ordered stream. The other stream consists of random sounds. There is always one higher and one lower stream and half of each stream is presented to the other ear.


**Fig. 3:** Structure of a single trial. The Target consists of two repetitions of the target stream alone (either the low or the high stream of notes). After that, the randomized stream joins it. The streaming can build up over 3 iterations of the basic sequence, before finally, the deviant occurs in one of the three last repetitions, A, B, or C.

Before conducting the test, participants underwent a training session. This training session was almost equal to the experiment itself, with the difference that the participants were given feedback whether they answered correctly. It was furthermore possible to repeat the stimuli.

The experiment featured further three conditions: a test condition for binaural fusion and two control conditions to verify that the task was indeed performed by fusing the target melody from the binaural input. Each of the three conditions was repeated 24 times.

### Binaural test condition

In the binaural test condition, both the target and the distractor stream were presented in the same fashion as the original scale illusion, i.e., every second note from each stream was presented to the other ear. This condition thus requires binaural fusion to perceive the two streams.

### Monaural control condition

In the monaural control condition, the target and the distractor streams were presented to the same ear. In this condition, listeners should be able to segregate the two streams based on their frequency range, even if they could not perform this segregation binaurally.

### Binaural control condition

In the binaural control condition, target and distractor streams were again presented binaurally, as in the binaural test condition – but binaural fusion was prevented by an altered timbre for all sounds in one ear.

In addition to pitch, also timbre represents a grouping cue (cf. Bregman, 1994; Deutsch, 1999). Here, it was altered severely by manipulating the temporal envelope and harmonics: Compared to the normal presentation, the duration of the notes itself have been shortened to 200 ms, while keeping the overall interval at 250 ms. Additionally, the half Hann-window ramps have been set to a duration of 50 ms. This corresponds to changes in the attack, release and sustain of the notes. Furthermore, harmonics of the pure tone at 2, 3, 4, and 5 multiples of the fundamental frequency have been added.

These changes were therefore expected to lead to a breakdown of performance that reflects the effect of binaural fusion when compared to the binaural test condition.

## ANALYSIS

A binomial distribution underlies this task, since the answer is either correct or false. Therefore, the significance levels for $p \approx 0.333$ are given by: $\Pr(X \geq 13, n=24) \approx 0.0284$ (* ; ≥ 54.17 %), $\Pr(X \geq 15, n=24) \approx 0.00323$ (** ; ≥ 62.50 %) and $\Pr(X \geq 16, n=24) \approx 0.00860$ (*** ; ≥ 66.67 %).

## RESULTS

The results are presented in Fig. 4 with the detection performance plotted as percent correct for the three conditions, binaural, binaural control and monaural. The average performance for the binaural and monaural conditions lies at about 80% correct (***), while the binaural control with altered timbre shows an average performance slightly above 50 % and still not significantly above chance performance.

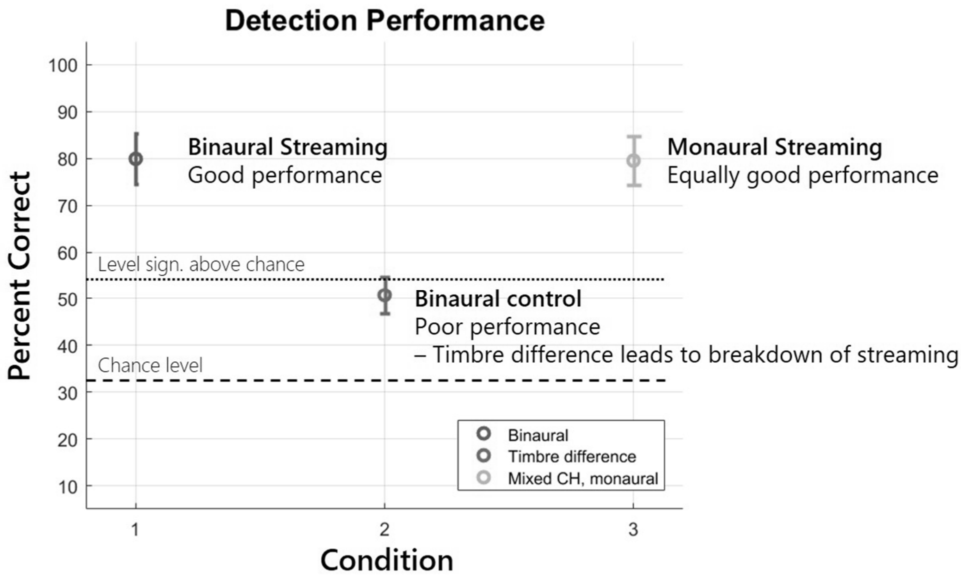


**Fig. 4:** Results of the scale illusion detection experiment in percent correct for the three conditions: **C1:** binaural streaming, **C2:** binaural control and **C3:** monaural streaming. The bars give the range for the standard error, the dashed line the chance level at 33.33 % and the dotted line indicates the performance level significantly above chance, 54.17 % (cf. Analysis).

## CONCLUSIONS

This study describes a task that proves binaural streaming and fusion ability of normal hearing participants. It is based on two melody streams that form the scale illusion. The participants showed that they can build a fused stream out of the binaural input, follow this stream and successfully detect the deviant note embedded into it. Further, in this task their monaural performance equals their binaural performance. The participants therefore do not have more difficulty following the streams binaurally. Additionally, the results of the binaural control condition demonstrate that a severe difference in timbre across ears leads to a breakdown of performance. When the two ears have such a severely different timbre, the components of the streams can no longer be identified as belonging together and the binaural information is no longer fused into a segregated object. Thus, participants can no longer follow the melody stream and identify the deviant note. Besides, this breakdown in performance shows that single-ear listening is insufficient to score well in the binaural condition, validating the experiment's design.

Variants of this task can potentially be used to test fusion ability of cochlear implant users (both bilateral and bimodal), where it is unclear whether binaural fusion takes place. This will be investigated in further studies.

## ACKNOWLEDGEMENTS

The authors would like to thank everybody who participated in the experiment for their help. We could not have done it without you!

## REFERENCES

Bregman, A.S. (**1994**). *Auditory Scene Analysis. The Perceptual Organization of Sound*. MIT Press, ISBN: 9780262521956.

Deutsch, D. (**1975**). "Two-channel listening to musical scales," J. Acoust. Soc. Am., **57**, 1156-1160. doi: 10.1121/1.380573

Deutsch, D. (**1999**). "Grouping mechanisms in music," in *The Psychology of Music*. Academic Press, ISBN: 9780123814609, pp. 299-348.

Middlebrooks, J.C., Simon, J.Z., Popper, A.N., and Fay, R.R. (**2017**). *The Auditory System at the Cocktail Party*. Springer Handbook of Auditory Research, **60**, doi: 10.1007/978-3-319-51662-2

Wickens, T.D. (**2002**). *Elementary Signal Detection Theory*. Oxford University Press, Oxford, ISBN: 978-0195092509, pp. 3-37, 93-110, 237-241.