



A modification of the scale illusion into a detection task for assessment of binaural streaming

Janssen, Niclas Alexander; Marozeau, Jeremy; Bramsløw, Lars; Riis, Søren Kamaric

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
[10.1121/1.5110243](https://doi.org/10.1121/1.5110243)

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Janssen, N. A., Marozeau, J., Bramsløw, L., & Riis, S. K. (2019). A modification of the scale illusion into a detection task for assessment of binaural streaming. *Journal of the Acoustical Society of America*, 145(6), EL457-EL462. <https://doi.org/10.1121/1.5110243>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A modification of the scale illusion into a detection task for assessment of binaural streaming

Niclas A. Janssen, Jeremy Marozeau, Lars Bramsløw, and Søren Riis

Citation: *The Journal of the Acoustical Society of America* **145**, EL457 (2019); doi: 10.1121/1.5110243

View online: <https://doi.org/10.1121/1.5110243>

View Table of Contents: <https://asa.scitation.org/toc/jas/145/6>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Structured phonetic variation facilitates talker identification](#)

The Journal of the Acoustical Society of America **145**, EL469 (2019); <https://doi.org/10.1121/1.5100166>

[A three-dimensional underwater sound propagation model for offshore wind farm noise prediction](#)

The Journal of the Acoustical Society of America **145**, EL335 (2019); <https://doi.org/10.1121/1.5099560>

[Revisiting the detection of interaural time differences in listeners with hearing loss](#)

The Journal of the Acoustical Society of America **145**, EL508 (2019); <https://doi.org/10.1121/1.5111065>

[Geoacoustic inversion with generalized additive models](#)

The Journal of the Acoustical Society of America **145**, EL463 (2019); <https://doi.org/10.1121/1.5110244>

[University student musician noise-dosage study measuring both ensemble and full-day noise exposure](#)

The Journal of the Acoustical Society of America **145**, EL494 (2019); <https://doi.org/10.1121/1.5110238>

[Highly spectrally undersampled vowels can be classified by machines without supervision](#)

The Journal of the Acoustical Society of America **146**, EL1 (2019); <https://doi.org/10.1121/1.5111154>



CAPTURE WHAT'S POSSIBLE
WITH OUR NEW PUBLISHING ACADEMY RESOURCES

Learn more 

AIP
Publishing



A modification of the scale illusion into a detection task for assessment of binaural streaming

Niclas A. Janssen^{a)} and Jeremy Marozeau

*Department of Health Technology, Technical University of Denmark, Ørstedsgade
Building 352, 2800 Kgs. Lyngby, Denmark
najan@dtu.dk, jemaro@dtu.dk*

Lars Bramsløw

*Eriksholm Research Centre, Rortangvej 20, 3070 Snekkerten, Denmark
labw@eriksholm.com*

Søren Riis

*Oticon Medical AIS, Kongebakken 9, 2765 Smørum, Denmark
SRKA@oticonmedical.com*

Abstract: Binaural streaming by frequency-proximity was investigated without subjective listener-feedback by modifying the scale illusion of Deutsch [J. Acoust. Soc. Am. **57**, 1156–1160 (1975)] into a detection-task. Nineteen listeners had to detect one deviation within a repeating melody stream, while simultaneously presented with a randomized distractor stream. Every second note in each stream was presented to the opposite ear, requiring binaural streaming to detect the deviant. Listeners performed well in this test but adding interaural delay or timbre-difference let the listeners group by lateralization instead. This confirms the grouping by frequency-proximity. The method could be used to investigate binaural streaming in hearing-impaired patients, where interaural percepts might differ.

© 2019 Acoustical Society of America
[BLM]

Date Received: February 18, 2019 **Date Accepted:** May 13, 2019

1. Introduction

In 1975, Deutsch aimed to investigate streaming of dichotically presented melodies and described an auditory percept called the scale illusion (Deutsch, 1975). The illusion was composed of two melody streams at different frequency ranges, one high and one low, out of the components of one musical scale. Both streams ascend and descend in frequency over eight notes. The streams were presented in such a way that every second note from each of the two streams was played to the other ear, in a pattern symmetrical to the middle of the repeating sequence of eight notes. So, when one component of the higher melody was in one ear, a component of the lower melody was in the other ear, and vice versa. The intriguing effect of this stimulation was that listeners did not group the stimuli by ear of input, but instead by frequency range. Most right-handed listeners perceived the higher stream coming from the right and the lower stream coming from the left, whereas only about half of the left-handed listeners did report this percept (“both streams”). A minority of the right-handed listeners and the other half of the left-handed listeners reported to perceive only the higher stream with four tones going up and down, and little or nothing of the lower stream (“single stream”). Deutsch argued that this illusion outlined binaural streaming in which the Gestalt-principle of pitch-proximity overrides the lateralization cues. The illusion is only possible through binaural processing of the stimuli from both ears.

The participants in the original study had to report verbally what they perceived. Describing the complex percept with two melodies precisely might be difficult for listeners without musical training and, likewise, the researcher might find it challenging to interpret the listener’s descriptions correctly. This study aims to validate a method in which the ability to perceive the scale illusion will be assessed through a detection task. Once validated this method could be performed by listeners with asymmetrical hearing loss and fitted with different hearing devices such as one hearing aid and one cochlear implant (CI), to evaluate their ability to integrate signals across ears and devices.

^{a)} Author to whom correspondence should be addressed.

2. Participants

Nineteen normal-hearing (NH) listeners participated in experiment I (aged 24 to 31; means 26.1 years, hearing-loss of less than 20 dB). Seven of the listeners were female, 11 musically trained, and three left-handed. They were all recruited at the Technical University of Denmark and provided written informed consent. All experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391).

3. Method

3.1 Stimuli

Participants were presented with sequences of notes that formed two melodies. Two types of stimuli were used to generate these melodies: pure tones and complex tones designed to vary on two distinct dimensions of timbre, impulsiveness, and brightness, as well as loudness. The interval between notes was 250 ms, as in the study by Deutsch (1975). All the pure-tone stimuli were loudness-balanced for each frequency and presented in a double-walled sound-isolated listening booth via Sennheiser HDA-200 headphones (Sennheiser electronic GmbH & Co. KG, Wedemark, Germany), driven by a Scarlett 2i2 soundcard (Focusrite Audio Engineering Limited, High Wycombe, UK).

The pure tones had duration of 250 ms with 10 ms half Hann-window ramps on the onset and offset. Their frequencies ranged from 262 to 523 Hz and were set according to the 12-tone equal-temperament tuning system. The complex tones were five-tone-harmonic complexes with duration of 200 ms including a 50 ms half Hann-window onset and offset. Their fundamental frequencies (F_0) were set to equal those of the pure-tones, and to this F_0 , the first four harmonics were added. All complex tones' components were presented 6 dB below the level of the related pure tone.

All the pure-tone stimuli were loudness-balanced using an adjustment procedure based on a reference sound: a pure tone at 508 Hz in the right ear, a frequency different from that of the pure-tone stimuli to be balanced. In the beginning, participants could adjust the reference sound to their preferred, most-comfortable loudness level. After that, they were presented with the reference sound followed by one of the pure tones in each trial. Their task was to adjust the loudness of the latter so that it matched that of the reference via an on-screen slider. The slider provided an adjustment range from 40 to 90 dB sound pressure level (SPL). Participants could repeat the stimuli and readjust the loudness until they were satisfied. This balancing was performed twice for all frequencies used in the experiment in random order. Overall, the participants chose levels that ranged from 56.3 to 85.3 dB SPL with a mean of 67.7 dB SPL. The headphones were calibrated using a Brüel & Kjær (B&K, Brüel & Kjær, Nærum, Denmark) 2636 sound level meter, with IEC-60318-1 ear simulator B&K 4153 and reference calibrator B&K 4230 and equalized by their frequency response.

3.2 Procedure

To test whether participants stream by frequency range instead of ear of input, this study used a detection-task paradigm, based on the scale illusion experiment by Deutsch (1975). As in the original study, two melodies were presented concurrently. These will be referred to as the target stream and the distractor stream. The target stream was designed 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 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 stream was of the lower frequency range, and vice versa (see Fig. 1). As in the experiment by Deutsch (1975), two notes were always presented simultaneously, one from each stream, except in a control condition with added delay (see Sec. 3.3). If a note from the target stream was on the right, a note from the distractor stream was on the left, and vice versa. Consecutive notes for both streams were presented to the opposite ear, in a pattern that reversed in order after every four notes. Thus, the arrangement formed a pattern symmetric to the middle of the repeating sequence of eight notes (e.g., for one stream in one presentation of the eight notes: left-right-left-right-right-left-right-left).

If the listeners group by frequency proximity, they should be able to follow the target stream in one ear, while the distractor stream is either perceived on the other side or not at all, as described by Deutsch (1975). Conversely, if the listeners were grouping based on lateralization cues, they would instead perceive random melodies on both sides, since the target notes and the random distractors from each side will be grouped. To test whether the listener could follow the target stream, a deviant note

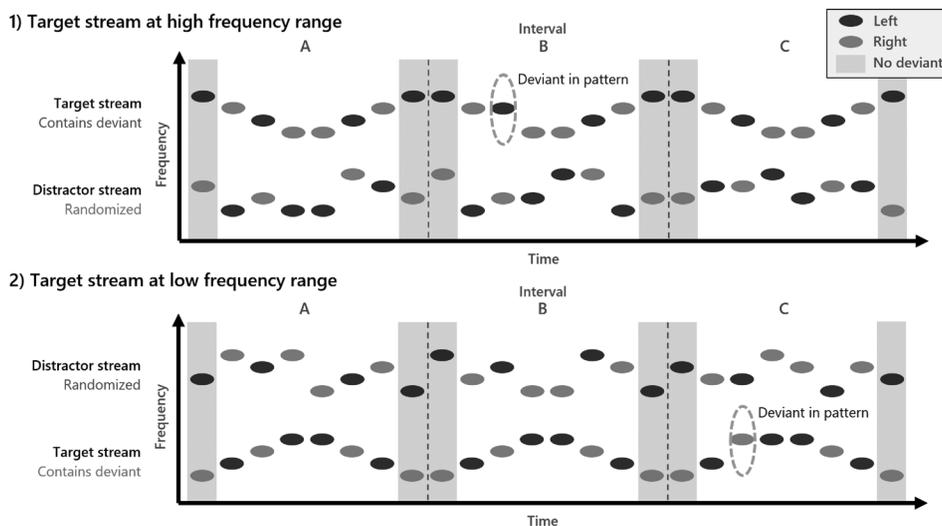


Fig. 1. Schematic stimuli representation, based on Deutsch (1975). A random deviant occurs in the target-stream, either at the higher (top) or lower (bottom) frequency-range. The distractor-stream consists of notes picked randomly from the other frequency range. The black and gray color of the ovals indicates the ear of presentation in the binaural test condition.

was introduced into it. The deviant consisted of shifting the note's respective frequency to one adjacent in the regular pattern. A listener who groups the binaural input by frequency-proximity was expected to detect this deviant easily, whereas somebody who groups by ear of input should experience difficulties to reliably detect the deviant note due to the unpredictable input in both ears.

At the beginning of each trial, the participant was presented with the target stream alone twice, indicating which melody to listen for. After that, target and distractor streams were presented simultaneously six times. Of these six repetitions, the first three were intended to allow for a build-up of streaming, since it has been reported that a segregated percept arises gradually over several seconds (Bregman 1990). One deviant note was introduced randomly in one of the last three repetitions, as depicted exemplarily in interval B at the top of Fig. 1. The interval being presented was indicated on a graphical user interface by lighting up the respective part of the stimulus on the screen. It was possible for the listener to repeat the stimuli.

The deviant could occur at any of the inner six of the eight notes within a presentation of the melody, but not at the first or last position. Its occurrence was counterbalanced with respect to the side of the presentation, interval, and frequency range of the target stream. Using a three-alternative, forced-choice paradigm prevented a detection bias, since the participants knew that a deviant was always present in one of the three intervals as described in Wickens (2002), therefore allowing for an "objective" assessment of their grouping behavior.

Before conducting the test, participants underwent training. This training was only different from the experiment, in that the participants were given feedback on whether they answered correctly.

3.3 Conditions

The experiment featured four conditions, which are described below. Each of the four conditions consisted of 24 trials and used the pure-tone stimuli, except for a binaural control with changed timbre. The order of all trials from all conditions was randomized per participant to prevent it from affecting results.

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, required grouping by frequency-proximity to perceive (i.e., segregate) the target stream and detect the deviant.

In the monaural test condition, the target and the distractor streams were presented to the same ear. In this configuration, the listeners should easily group the target notes into one stream and the distractor notes in another one and, as a consequence, accurately detect the deviant. This condition was designed to verify that the listener could detect the deviant once the target notes were grouped into one stream.

As the ability to detect the deviant is used as a proxy to evaluate whether the listener can experience the scale illusion, it is necessary to add a control condition in which grouping the target notes across ears will be prevented by enhancing the lateralization cues to induce grouping by ear-of-input. In this configuration, the listener should perceive two random melodies in each ear, and should not be able to detect the deviant.

Two binaural control conditions were tested, in which the lateralization cues were enhanced through two different methods. In the first control condition, the binaural delay control condition, a delay of 125 ms (half a note's duration and larger than the range for natural interaural time-difference cues) was added to the sounds in one ear. As demonstrated by Deutsch (1979), introducing an asynchrony increases the tendency to treat each ear-input as emanating from a different source and will, therefore, induce a grouping by ear-of-input. Only eight of the 19 participants were tested in this condition.

In the second binaural control condition, a significant timbre and loudness difference was introduced between the notes presented in each ear. Pure tones were presented to one ear as in the other conditions, whereas complex tones (cf. Sec. 3.1) were presented to the other. Adding a salient perceptual difference across ears should lead the listener to group the notes by ear-of-input and result in a low detection-performance. This condition will also help to quantify the possible performance when attending to just one ear.

4. Results

The results are presented in Fig. 2. The detection performance is plotted as a percent of correct answers for the four conditions: binaural and monaural test conditions, followed by the two binaural control conditions. For each condition, there are two data points: the average performance for the target stream in the low frequency-range (“low”), and the average performance for the target stream in the high frequency-range (“high”). The average performance for both the binaural and monaural test conditions lies at about 80% correct, and the performance in the high frequency range is slightly higher than in the low frequency range. The average performance in the binaural control condition with delay is much lower at about 54% correct, while that in the control condition with altered timbre lies at around 50% correct. In both controls, again the performance in the high frequency range is slightly higher. The trend for high performance in both test conditions and lower performance in both controls was also found for each listener individually.

A two-way repeated analysis of variance was performed in the software JMP[®] (SAS, Cary, NC) with the “rationalized” arcsine transform applied (rau; Studebaker

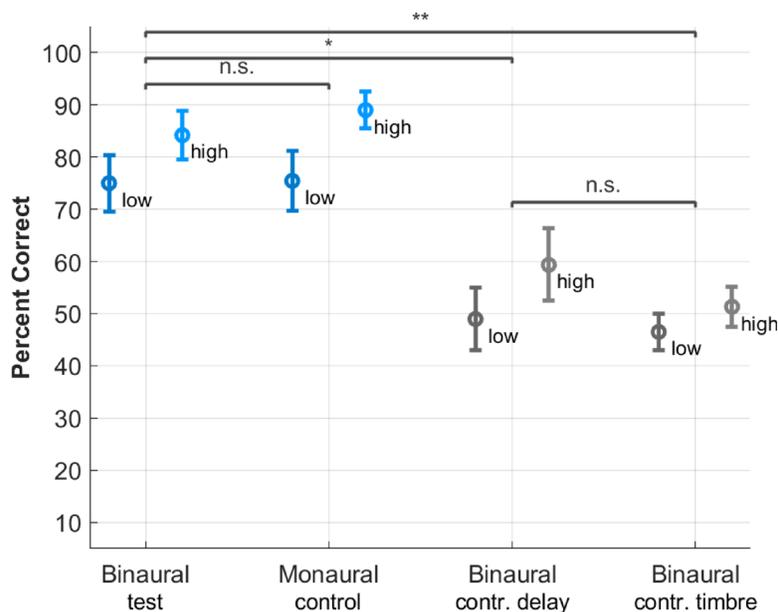


Fig. 2. (Color online) Results from 19 NH-listeners in percent of correct answers for the four conditions. Only eight of them were tested in the binaural delay control condition. In each condition, the average performance is plotted for the target-stream in the high (high), and low (low) frequency-ranges with the standard error and indication of significant differences.

1985), detection scores as the dependent variable, and the conditions, as well as the target stream's frequency range as factors. Both the condition [$F(3, 43.9) = 37.0$, $p < 0.0001$], as well as the frequency range [$F(1, 19.6) = 5.76$, $p = 0.0264$] are significant factors, while their interaction is non-significant [$F(3, 44.29) = 1.11$, $p = 0.355$]. The estimated degrees-of-freedom are decimals because the number of participants tested in the delay-condition is lower than in the other conditions. *Post hoc*-analysis using the Bonferroni correction and again *rau*-transformed data, shows that results for both the binaural and monaural test-conditions are not significantly different from each other ($p = 0.118$). The results in the binaural test-condition are significantly different from the binaural controls with delay ($p = 0.0138$) and timbre ($p = 0.0012$), while results in these two binaural controls are not significantly different from each other ($p = 0.174$).

5. Discussion

The main aim of this study was to develop and validate a method in which the ability to perceive the scale illusion was assessed through a detection task. Three binaural and one monaural conditions were tested. We assumed that if the listeners can experience the scale illusion, they should be able to detect a deviant in the binaural test condition, but not in the two binaural control conditions. A monaural test condition was added to evaluate the average detection score that can be obtained when the target notes were easily grouped. The average score obtained shows a moderately high and similar score for both the binaural and monaural test conditions (see Fig. 2). Furthermore, both control conditions yielded a score significantly lower than the binaural test condition. Taken together, these results indicate that this method can be used to demonstrate the ability to experience the scale illusion without the need to describe the melody pattern as in the original study (Deutsch, 1975).

Recently, Metha *et al.* (2016, 2017) have used the octave illusion (Deutsch, 1974) to assess binaural streaming in NH listeners, using it as the base for a detection-task with either loudness or modulation cues. The octave illusion was based on only two tones, one high and one low and separated in frequency by an octave (400 and 800 Hz). These tones were presented in a dichotic sequence, so that one ear received a pattern high–low–high–low, while the other received the inverse pattern, low–high–low–high, simultaneously. As with the scale illusion, listeners could group these stimuli either based on pitch or based on location. The most common percept was a high tone lateralized to one ear alternating with a low tone lateralized to the other. Metha *et al.* (2016) have modified the original method into a detection task, in which the participants had to detect a variation of loudness in a target stream. Their results demonstrated that the illusion is subject to the same constraints as auditory stream segregation and how the listener's attention influenced the obtained percepts, i.e., their lateralization. Further results in Mehta *et al.* (2017) revealed that the octave illusion arises from a "misattribution of time across perceptual streams, rather than a misattribution of location within a stream," providing a better understanding of the mechanisms involved in binaural streaming.

In the original experiment (Deutsch, 1975), the participants report hearing either one or two melodies. It is impossible to tell just from the performance in the binaural test condition whether listeners perceived one stream or both high and low streams simultaneously, but with their high detection-scores for the target in either frequency-range, they must at least be able to attend to either stream selectively. However, the frequency range in which the deviant note is embedded had a significant impact on performance. Consistent across conditions, the detection performance in the higher frequency-range is slightly higher, even though the stimuli were loudness-balanced at the different frequencies. Therefore, participants were able to focus selectively on either the low or high stream, but the better performance in the higher frequency-range suggests that focusing on the higher stream was easier. This could explain the description of the single stream percept in Deutsch (1975), where listeners reported to perceive only the higher stream and nothing or little of the lower one. If it is easier to focus on the high stream and listeners voluntarily are not given further instructions, they should then be more likely to focus purely on the higher stream. The reported single stream percept is thus not at odds with the current results. Furthermore, these results match one of the conclusions from Deutsch (1985), where musically-trained listeners had to identify dichotically presented sequences of tones correctly. There, listeners also performed consistently higher in transcribing higher tones compared to lower tones.

The effect of timbre on the streaming behavior could have implications for hearing-impaired patients, depending on the aids they use for listening. If these aids do

affect the timbre of sounds differently across ears, binaural streaming could be affected negatively. This could apply to patients with CIs, providing electric instead of acoustic stimulation, especially for patients with a unilateral CI and acoustic hearing on the other side and this task could be adapted to assess if these changes in interaural percepts lead to changes in binaural streaming behavior for CI patients. Also, patients with highly asymmetric hearing loss, using entirely different hearing-aids across ears, could be affected. If the binaural object-formation was malfunctioning, performance in tasks relying on the evaluation of binaural cues, such as speech reception in noise and sound localization, could be reduced severely.

6. Conclusion

This study proposes a modification of the scale illusion experiment by Deutsch (1975) into a detection-task to assess listeners' binaural streaming behavior without relying upon subjective reports of percepts. As in the original study, the NH listeners grouped the binaural stimuli by frequency proximity instead of lateralization, for which they had to integrate stimuli from both ears into a common stream. Participants were able to focus on either the low or high stream and detect a deviant note, but it appears to be easier to focus on the higher stream. Two control conditions were added in which lateralization cues were enhanced. Detection scores dropped significantly, indicating that the task was facilitated by the ability to group stimuli by frequency proximity. The new task could be suited to assess binaural streaming in the hearing-impaired population and explore possible changes and adaptations in the mechanisms.

Acknowledgments

The authors would like to thank everybody who participated in the experiment for their help and Tanmayee Uday Pathre for collecting part of the data. Thanks to the colleagues from the Hearing Systems Group for valuable comments and discussions, and Andrew Oxenham for a fruitful discussion in an early stage of the experiment. These studies were made possible by funding from the Centre for Excellence in Hearing and Speech Sciences (CHeSS), Oticon Medical, and the Oticon Foundation.

References and links

- Bregman, A. S. (1990). *Auditory Scene Analysis. The Perceptual Organization of Sound* (MIT Press, Boston, MA).
- Deutsch, D. (1974). "An auditory illusion," *Nature* **251**, 307–309.
- Deutsch, D. (1975). "Two-channel listening to musical scales," *J. Acoust. Soc. Am.* **57**, 1156–1160.
- Deutsch, D. (1979). "Binaural integration of melodic patterns," *Percept. Psychophys.* **25**(5), 399–405.
- Deutsch, D. (1985). "Dichotic listening to melodic patterns and its relationship to hemispheric specialization of function," *Music Percept.* **3**(2), 127–154.
- Mehta, A. H., Jacoby, N., Yasin, I., Oxenham, A. J., and Shamma, S. A. (2017). "An auditory illusion reveals the role of streaming in the temporal misallocation of perceptual objects," *Philosoph. Transact. R. Soc. B* **372**(1714), 20160114.
- Mehta, A. H., Yasin, I., Oxenham, A. J., and Shamma, S. (2016). "Neural correlates of attention and streaming in a perceptually multistable auditory illusion," *J. Acoust. Soc. Am.* **140**(4), 2225–2233.
- Studebaker, G. A. (1985). "A 'rationalized' arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Wickens, T. D. (2002). *Elementary Signal Detection Theory* (Oxford University Press, Oxford), pp. 3–37, 93–110, and 237–241.