

# Detection and Recognition Threshold of Sound Sources in Noise

Carina Pals (c.pals@ai.rug.nl)

Auditory Cognition Group, Department of Artificial Intelligence, University of Groningen  
Bernoulliborg, Nijenborgh 9, 9747 AG Groningen, the Netherlands

## Abstract

This study examines detection and recognition thresholds for environmental sounds in the presence of noise. Human listeners were presented with a selection of everyday sounds masked by noise at different signal-to-noise ratios (SNR). Participants had to indicate if they detected or recognized the sound in the masking noise. Different categorizations of the sounds were compared. Results show that pulse-like sounds are detected at a lower maximum local SNR than noise-like sounds. For one categorization we found detection of tonal sounds at a lower maximum local SNR than noise-like sounds. For another categorization we found detection of pulse-like sounds at a lower maximum local SNR than tonal sounds. These differences in detection between pulse-like, tonal and noise-like sounds suggest human auditory perception combines different strategies for detecting these sound types.

**Keywords:** Auditory perception; environmental sound; sound source recognition.

## Introduction

This study will take a closer look at the human ability to detect and recognize common sound sources. Humans have evolved to detect and recognize sound sources in order to interpret and act on events in their environment. Surprisingly, little research has been aimed at understanding this ability to recognize environmental sounds. Some interesting studies on recognition of environmental sounds are Gaver (1993), Ballas (1993), and Gygi, Kidd, & Watson (2007). The more popular areas of study in auditory perception have been music perception, speech recognition, and the perception of idealized signals such as tones and pulses as studied in psychoacoustics. Research on human sound source recognition can augment this existing research to provide a broader and more integrated understanding of auditory perception. Apart from its value to fundamental knowledge and research on perception, it can also be used for more practical purposes. For example, human performance data on sound recognition tasks can be used to determine how well an automatic sound recognition system approaches human performance.

A theory of human sound source recognition can aid the further development and refinement of automatic sound recognition systems. New insights will provide new strategies for identifying signal components and assigning them to the correct sources. This facilitates the separation of a signal into sounds from different sources, for example speech from non-speech. To improve automatic sound recognition, it is important to know which features of environmental sounds are robust to signal degradation. In the case of speech recognition, voiced parts of speech have been shown to be more robust than unvoiced (aperiodic) parts of speech. Though voiced speech is

typically louder than unvoiced speech, other properties of voiced speech such as harmonicity contribute to this robustness. Identifying robust features of environmental sounds will help develop noise robust systems.

This experiment attempts to find robust features for sound recognition by measuring the detection and recognition thresholds of a variety of common sounds in masking noise. The experimental sounds were chosen using recent research by Gygi et al. (2007), which divided everyday sounds into three categories based on similarity ratings for pairs of sounds as estimated by the participants. Since similarity judgments are likely to be based on the more prominent features of the sounds, which in turn are more likely to be robust to noise, we argue that the acoustical features that distinguish these categories might include the robust features for environmental sound recognition. This warrants a closer look at these categories. Our experiment focuses on differences in detection and recognition thresholds between sounds from Gygi's categories, to determine which category contains the more robust sounds. However, the classification method Gygi et al. use for categorizing sounds does not provide definite indications of the underlying acoustical features. To gain better insight into what acoustical properties may lead to robustness, we also compare our results to a model-based categorization based on Andringa (2008).

The next section explains the work by Gygi et al. and Andringa in greater detail, as well as some other relevant experimental and theoretical work. The following section describes the design and method of our experiment, followed by a presentation of the results. The last two sections present a discussion of the results, some recommendations for follow-up studies, and our conclusion.

## Theoretical Background

### Categories of Common Sounds

Recent research by Gygi et al. (2007) divides a broad range of sounds into three distinct categories: harmonic sounds, discrete impact sounds, and continuous sounds. These categories were based on pairwise comparisons of 100 different sounds by four participants. This led to a perceptual distance matrix of 100 x 100 item-item distances. The Gygi categories are derived from a linear separation of the first two dimensions of a multidimensional scaling (MDS) solution of this perceptual distance matrix (see figure 1). MDS places items in a low dimensional space with distances between the items that approximate the original set of distances. The first MDS dimension explains most of the variance in



As far as we know, no integration time-constant for pulse-like signals has been estimated. In fact the diverse thresholds described in this paper can help to identify the strategies and associated parameters for (pulse-like) environmental sounds. In this study we use the model described in Andringa (2008) with frequency-channel dependent integration time-constant equal to two times the channels best period (the inverse of the channel center frequency) with a minimum of 5 ms (for all center frequencies above 500 Hz).

### Identification of Common Sounds

Ballas (1993) identifies a number of factors that have an effect on the identification of short everyday sounds. These factors include ecological frequency, familiarity, causal uncertainty, and typicality of the sound. Ecological frequency and familiarity are closely related; ecological frequency refers to how often a sound occurs in the natural environment, the more often a sound occurs, the more familiar it will be. Ballas defines causal uncertainty as uncertainty about the source of the sound as an effect of the number of sources that can produce similar sounds. Typicality of the sound is explained as how closely a sound resembles a mental representation of a stereotypical sound produced by the type of sound source. These factors are very important when choosing sounds for studying sound recognition in human listeners.

Another factor playing a role in the identification of sounds in an experimental setting is prior knowledge. With no prior knowledge a noise degraded sound will only be recognized when enough information is present in the noisy signal to yield the right hypothesis. With prior knowledge about the set of sounds in the experiment participants can search for evidence in the noise for the presence of this limited number of sounds. This results in a lower threshold for recognition with prior knowledge of the sounds used in the experiment. Therefore we expect to find a higher recognition threshold for sounds in the first half of the experiment when going from noisy to clean signals, and a lower recognition threshold in the second half of the experiment, when going from clean to noisy. This effect is similar to hysteresis, though strictly speaking hysteresis is a difference in thresholds depending on direction only. In this case it is a difference depending on states; without prior knowledge, or with prior knowledge.

## Method

### Participants

Twelve students of the University of Groningen with average hearing participated in the final version of the experiment as an obligatory part of a perception course. One participant reported being 50% deaf in one ear and 80% in the other, these data were excluded from the analysis.

### Stimuli

This experiment uses sound from the set used by Gygi et al. (2007) in their study of similarity and categorization of environmental sounds. A total of thirty sounds were selected from their set of 100 sounds, ten from each of the three categories shown in figure 1. Each set of ten was chosen so that half of the sounds are close together in the center of the cluster and the other half are evenly distributed over the areas towards the borders between the categories. Criteria for selection also included an initial estimation of the recognizability of the sounds keeping in mind the effects described by Ballas (1993), to reduce bias caused by differences in ease of recognition.

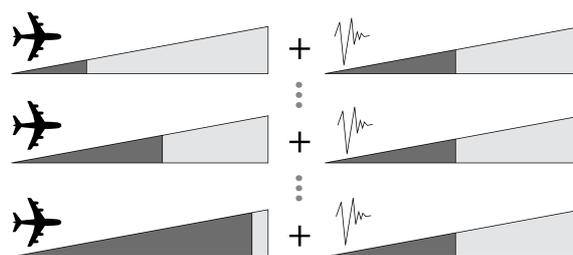


Figure 3: noisy stimuli for target sound airplane

The selected target sounds were used to create noisy stimuli for use in the experiment. The average loudness of the noise used to mask the target sounds was the same for all noisy stimuli. The loudness of the target sound was adjusted for each noisy stimulus to reach the desired maximum local SNR. A schematic representation of the creation of noisy stimuli is shown in figure 3. This method was chosen to prevent the loudness of the noise from becoming a predictor for the target sound or sound type. Additional measures were taken to further minimize possible predictors for target sounds. All sounds are four seconds long, to prevent signal duration acting as a predictor. The point in time at which the signal is introduced varies per noise level as well as per target sound. Each noisy stimulus was created with a different subset of pink noise, to avoid any local loudness patterns in the noise itself becoming a predictor for the target sound or sound type.

To create the noisy stimuli, the maximum local SNR was calculated as follows. For the noise segment as well as the target sound a cochleogram was generated. A cochleogram is an auditory perception inspired spectrogram-like time-frequency representation (Andringa 2002, Andringa 2008) in the form of a matrix containing the signals energy levels in dB per time-frequency point, which can be visualized as shown in figure 4. The matrix containing the energy levels of the noise segment was then subtracted from the energy matrix of the target signal, resulting in a matrix of local signal-to-noise ratios for each time-frequency point. We refer to the maximum value in this local SNR matrix as the maximum local SNR of the signal. Each target sound was amplified to achieve the desired maximum local SNR before adding the target and noise signals together to create the noisy stimuli.

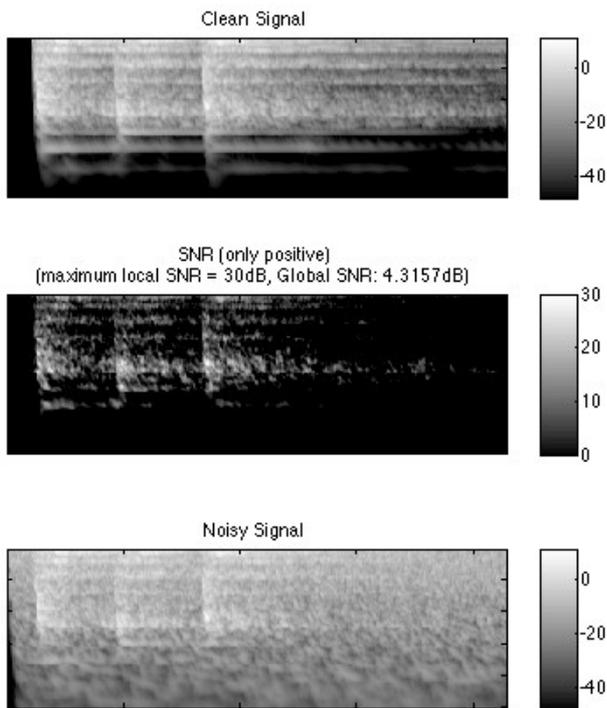


Figure 4: Example cochleograms for cymbals

For each of the 30 target sounds, a set of 22 noisy stimuli was created, each with a different maximum local SNR. The 22 maximum local SNR values to be used in the experiment were first chosen per target sound by listening to each noise masked stimulus and estimating the detectability and recognizability. The resulting set of 22 maximum local SNR values differs for each target sound. During a pilot preceding the experiment, the range of SNR values for each target sound to be used in the experiment was adjusted where needed. The interval in maximum local SNR between the noisy stimuli was not constant throughout the whole range of 22 noisy stimuli. Around the estimated detection and recognition thresholds for each of the target sounds the intervals in SNR between the noisy stimuli were 1dB, while in less interesting regions the intervals were 3 or even 5 dB.

## Materials

The stimuli were all well above the ambient noise level. Closed-back headphones (Sennheiser HD 215) were used. The experiment was run on a laptop and presented to the participants using a graphical user interface built in Matlab.

## Procedure

The average time participants needed to complete the experiment was approximately 45 minutes. For each new run of the experiment 17 sounds were selected randomly from the total set of 30. Ten were used as target sounds in the experiment, six were used as filler sounds, and one was used in an example before the measurements started. At the start of the experiment participants were presented with a short description detailing how to interact with the

interface, followed by three example items. After these examples, the interface would proceed to the actual experiment and start recording data. The general flow of the experiment is illustrated in figure 5.

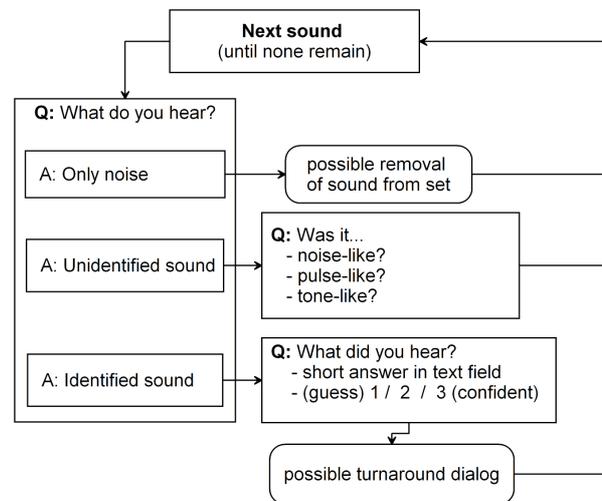


Figure 5: flow of the experiment

Items in the experiment consisted of a noisy stimulus being played, followed by a few questions. First the participant was expected to indicate what they heard: nothing but noise, something unrecognizable, or something they recognized. In case the participant reported to have heard nothing but noise, the experiment proceeded to the next item. In case the participant reported to have heard something unrecognizable, the next and final question of the item was to indicate whether the sound heard in the noise was noise-like, tonal, pulse-like, or a combination of the three. In case the participant answered to have recognized the sound in the noise, they were asked to name the source and give their answer a confidence score ranging from 1 (guess) to 3 (certain). After all questions were answered, the participant could press 'next' to proceed to the following item. All answers given by the participant were recorded.

When deciding which sound to present next, the first choice is between presenting something from the set of target sounds (70% chance) or filler sounds (30% chance), the next step is to randomly select a sound from the chosen set. This resulted in a continuous mix of sounds intended to avoid the effect of participants searching for evidence of a specific sound, which would be expected if all stimuli for a target sound were to be presented consecutively. Filler sounds serve this same purpose. The fillers also serve to ensure a less boring, more varied set of noise levels to avoid loss of attention that may happen at stages of the experiment when most presented target sounds are nearly undetectable. Filler stimuli were chosen randomly from the range of the noisy filler stimuli available, excluding those consisting of mostly noise, and the ones most easily recognizable.

Though the order in which the different target sounds were presented was random, the noisy stimuli for each specific target sound were presented in a fixed order. The

noisy stimuli were first presented in order of increasing maximum local SNR. After confident recognition by the participant, or after presentation of the noisy stimulus with the highest maximum local SNR, the stimuli were presented in order of decreasing maximum local SNR until they the target sound was no longer detected. The order in which the noisy stimuli were presented for a single target sound is illustrated in figure 6.

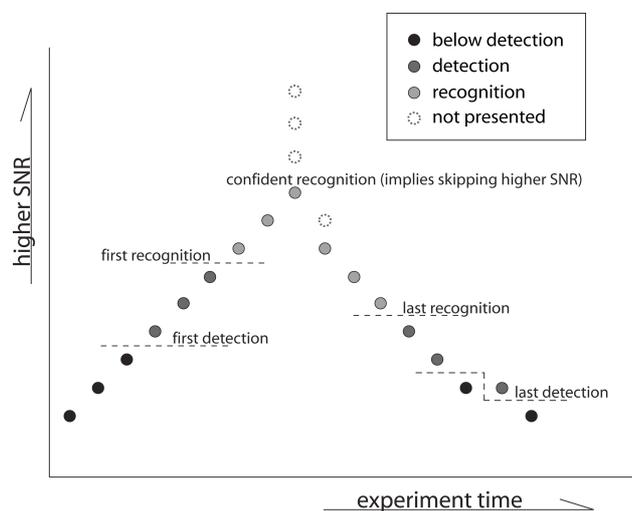


Figure 6: order of SNR levels presented during the experiment for one target sound

We expected initial detection and recognition to be at a higher maximum local SNR when the participants had no prior knowledge about the sounds used in the experiment. Therefore, for each target sound the five noisy stimuli with the lowest maximum local SNR were skipped at the start of the experiment. The noisy stimuli were first presented in order of increasing maximum local SNR. When a participant recognized a target sound with confidence (confidence rating 3), a popup dialog was displayed. This dialog stated the name of the sound source, allowed the participant to listen to a clean version of the target sound (without masking noise), and asked if the participant had recognized the sound correctly. This information was recorded but also verified afterwards.

The dialog playing the clean target sound was also presented if the participant reached the noisy stimulus with the highest maximum local SNR of the 22 noise masked versions of the target sound without confidently recognizing the sound source. This was done to make sure the participants knew the name of the sound source and had heard the clean signal. After presentation of the clean signal, regardless of whether this was because the sound was recognized with confidence, or because the signal with the highest maximum local SNR was reached, the noisy stimuli were presented in order of decreasing SNR. After the presentation of the clean signal one SNR level was skipped, the next noisy stimulus to be played was the one with the second-lower maximum local SNR.

When a target sound was no longer detected at a particular maximum local SNR level, the target sound was not immediately removed from the set. Instead, the next

time the target sound was chosen for presentation the noisy stimulus with same maximum local SNR level was presented again, in order to make sure the the participant consistently did not detect the target in the masking noise. In case the sound was detected this second time, the target sound remained in the set and the next lower maximum local SNR would be presented the next time the target sound was selected. The second time a participant reported not detecting a particular target sound, whether at the same or a lower maximum local SNR, that target sound was removed from the set. The target sound was also removed when the lowest maximum local SNR in the set was reached. The experiment terminated when the last of the target sounds was removed from the set.

## Analysis

The differences between categories were analyzed with Friedman's test. Friedman's test evaluates within-subject differences between three or more groups of data. For each categorization we tested whether the difference in maximum local SNR for first detection, first recognition, last recognition and last detection between the three categories are significant.

For each of the six categorizations we took the medians of the maximum local SNR at which each participant first detected, first recognized correctly, last recognized correctly and last detected the sounds in each of the categories. The resulting data we used to do a Friedman test for all six categorizations for each of those 4 conditions.

When a Friedman test results in significant differences it is necessary to compare the differences between all pairings of the categories to determine which differences are significant and which are not. Such a multiple comparison procedure gives the mean rankings for each category, the 95% confidence intervals for the mean rankings and information on which differences are significant and which are not.

## Results

The graph in figure 7 depicts the mean maximum local SNR and standard deviation for which each sound in the experiment was first detected and those for which it was last correctly recognized. The sounds are sorted by their mean maximum local SNR for first detection. Either side of the graph shows the names of the sounds, color coded to show what category they belong to for two of the categorizations. The left axis shows Gygi's categorization, and the right axis the categorization based on Andringa's algorithm for signals with a maximum local SNR of 6dB. It is important to note that seven sounds were unanimously detected at the first presentation. These sounds were the cymbals, clock, footsteps, breaking glass, typewriter, keyboard, and ping-pong ball.

For both categorizations the sounds appear to be grouped per category along the axis, though there is some overlap between the categories. Discrete impact / pulse-like sounds cluster around one end, harmonic / tonal sounds in the middle and continuous / noise-like around the other end. This suggests pulse-like sounds are detected

at a lower maximum local SNR than tonal sounds which in turn are detected at a lower maximum local SNR than noise-like sounds. We tested if these differences are indeed significant with Friedman's test.

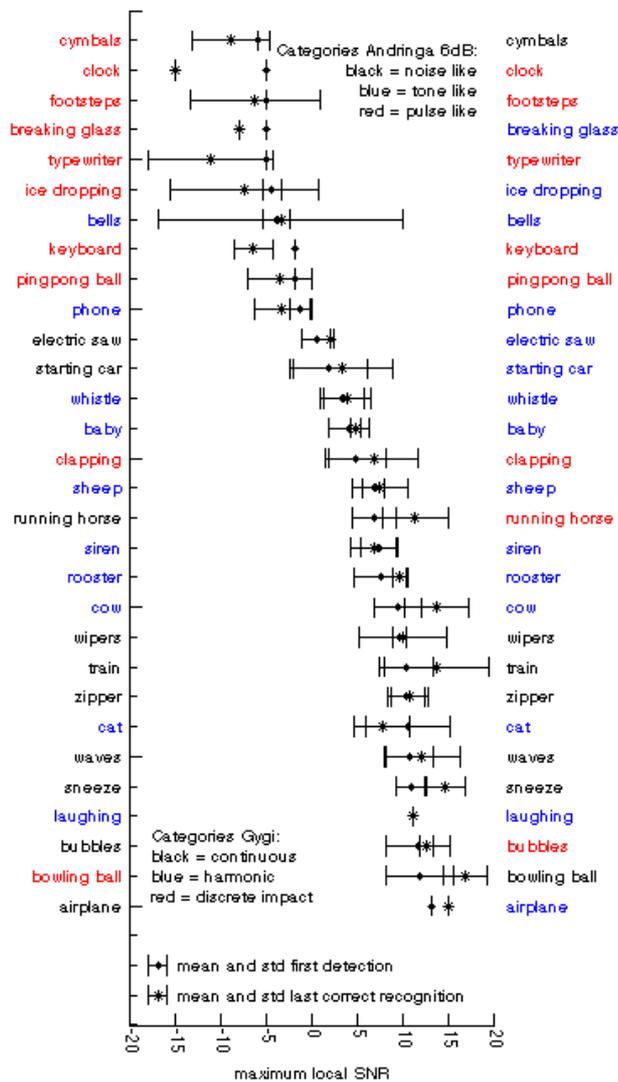


Figure 7: mean maximum local SNR values for first detection and last correct recognition

An interesting observation about the results of the experiment is that first detection seems to coincide with last correct recognition. Figure 8 shows the correlations and mean differences between the thresholds measured for first detection, first recognition, last recognition, and last detection. These correlations and mean absolute differences were calculated over the thresholds measured per participant per sound. The correlation between the first detection and last correct recognition is quite high (0.89). However, the correlation between first and last detection is even higher (0.93), as is the correlation between last detection and last correct recognition (0.94). The correlation between first correct recognition and the other three thresholds was relatively low. This can be

explained by the wide variance in the data for first correct recognition.

The smallest mean absolute difference is the one between last detection and last correct recognition. The difference measured between these two is roughly 2 dB. The difference between first detection and last detection as well as the difference between first detection and last correct recognition appears to be about 4 dB. It is interesting to note that the difference between first recognition and the other thresholds is approximately 10dB, which agrees with the differences in thresholds for detectability and intelligibility of speech in white noise found by Hawkins and Stevens (1950).

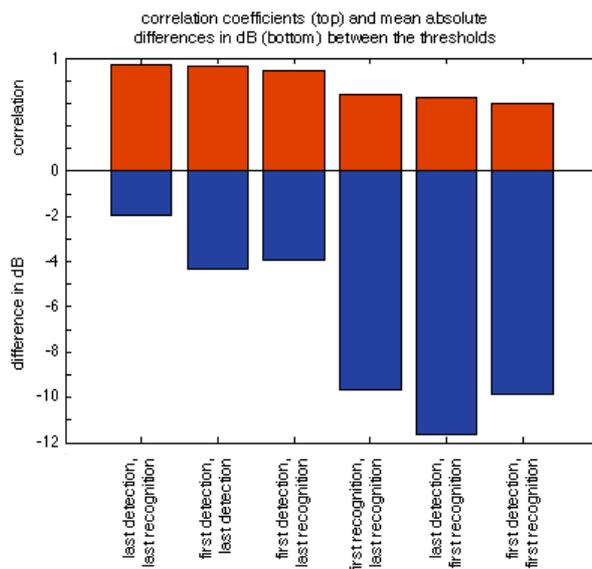


Figure 8: correlations and mean absolute differences

Figure 7 shows considerable overlap of the one-standard-deviation error bars between sounds. Importantly, though not visible in the plot, the number of listeners for each individual sound varies between 1 and 7 so the confidence level for the mean thresholds for some sounds is quite low. This means we can not draw specific conclusions about the mean maximum local SNR values for which the individual sounds are typically first detected, first correctly recognized, last correctly recognized, and last detected. However, if we group the sounds by category we can make inferences about differences between the categories. Knowledge about categories of sounds is important since it can be generalized to include novel sounds.

The results of the Friedman test, the analysis of variance by ranks is shown in figure 9. The graphs in this figure show the p-values for each categorization-condition pair. The six categorizations are the Gygi categories, and the five categorizations based on Andringa (2008) for the time-frequency regions of the sounds not masked by noise for maximum local SNR values of 3dB, 6dB, 9dB and 18dB, and for clean signals. These plots show that for all categorizations, the differences in detection or recognition between the categories is highly significant, with the exception of the differences in first correct recognition.

The differences in first recognition between the categories based on Andringa (2008) for maximum local SNR levels of 3dB and 9dB show a trend towards significance.

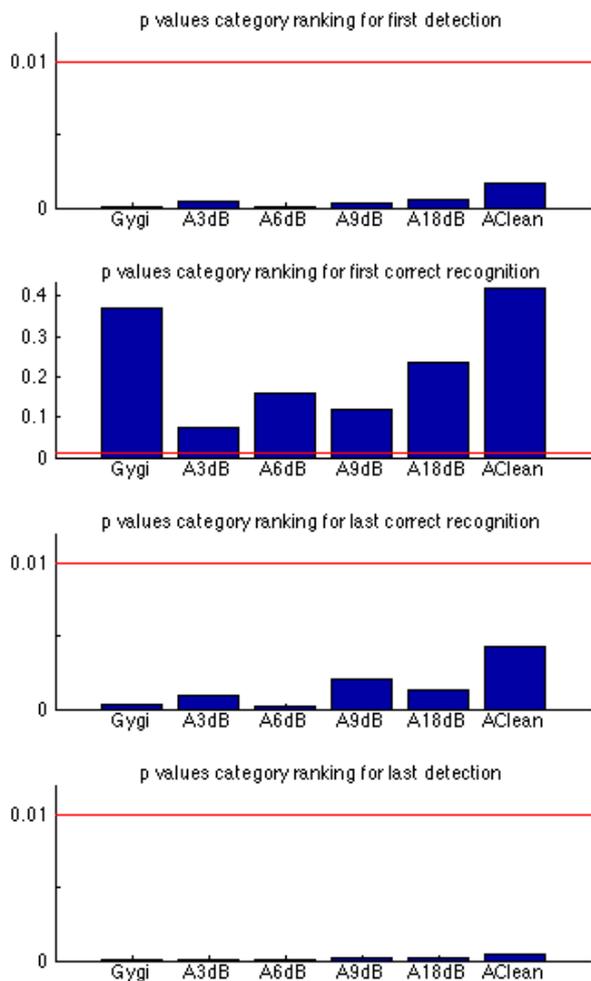


Figure 9: Friedman test p-values

The Friedman tests were followed by a multiple comparison procedure to see which differences between categories were significant on a 95% confidence interval and which were not. Some of the results of this comparison are shown in figure 10. As can be seen in this figure the bars representing the 95% confidence interval for the mean ranking of the category tonal sounds overlap with at least one other category in all figures and in some cases with both. This means the difference between tonal sounds and the category the bars show overlap with is not significant on a 95% significance level for that categorization.

The figure shows that for the Gygi categorization the difference in first detection between discrete impact sounds and both harmonic and continuous sounds is significant. However, the difference between harmonic and continuous sounds is not significant. For last correct recognition only the difference between discrete impact and continuous sounds is significant; the bar representing

the 95% confidence interval for the mean ranking of harmonic sounds overlaps with both the bars for discrete impact and continuous sounds. However, the means for last correct recognition for these three categories are distributed quite evenly, which suggests the categorization may provide a good indication of ranking for last correct recognition. This is also the case for categorizations based on Andringa (2008) for sounds with a 9dB and 18dB maximum local SNR. This warrants further investigation.

For the categorization based on Andringa (2008) for sounds with a 6dB maximum local SNR the difference between noise-like sounds and pulse-like as well as tonal sounds is significant. The bars representing the 95% confidence interval for pulse-like and tonal sounds however show quite some overlap.

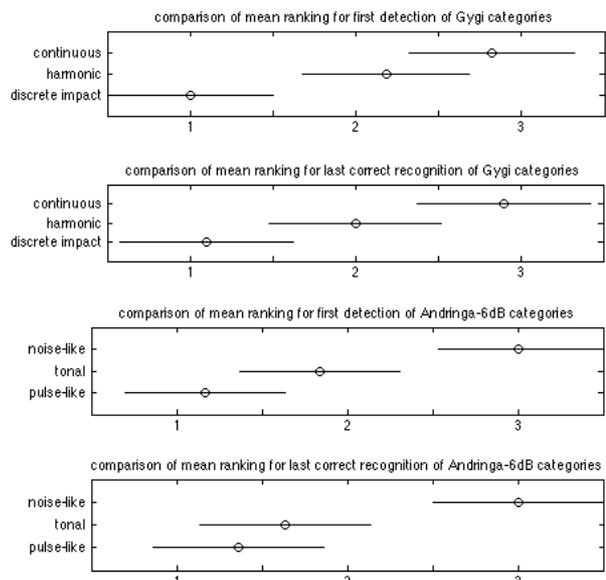


Figure 10: comparison of the categories mean ranking including 95% confidence intervals, no overlap in bars implies significant difference

For all categorizations the difference in ranking for first detection, last correct recognition, and last detection between pulse-like and noise-like sounds was significant. For first correct recognition, none of the differences in ranking were significant on a 95% confidence interval. The different categorizations based on Andringa (2008) differ in which category shows the most significant difference. The categorizations based on sounds masked by noise at a low maximum local SNR distinguish better between sounds categorized as noise-like and the other sounds. The categorizations based on clean signals and sounds with high maximum local SNR distinguish better between pulse-like sounds and the other sounds. Table 1 shows which category was significantly different from both other categories for all possible categorization-condition pairs.

Table 1: the most distinctive categories per categorization-condition pair

	first det.	first rec.	last rec.	last det.
Gygi	Pulse	-	-	Pulse
A3dB	Noise	-	Noise	Noise
A6dB	Noise	-	Noise	Noise
A9dB	-	-	Noise	-
A18dB	Pulse	-	-	-
AClean	Pulse	-	Pulse	Pulse

## Discussion

### Determining the SNR

Pulse-like sounds appear to be detected at a lower maximum local SNR than tonal sounds. One of the sounds, the ticking clock was last detected by all participants at the lowest SNR used in the experiment, a maximum local SNR of -15dB. The fact that some pulse-like sounds are still detected at such low maximum local SNR levels is most likely a consequence of the way the maximum local SNR was determined. As was explained earlier, the SNR is determined by integrating the signals energy over a small period of time. However, ideal pulses are characterized by an energy peak at one single point in time. An integration time window suitable for determining the SNR of tonal sounds is longer than the duration of a pulse. For pulse-like sounds this will result in a lower SNR than the actual ratio between energy levels at the time of the pulse.

The fact that human listeners detect pulses masked by noise at such a low computed local SNR suggests that the human auditory system uses a smaller integration time-constant. For pulse detection the auditory system must be capable of integrating spectral information, while it must integrate temporal information for tone perception. Perception of noise-like sounds depends on both spectral and temporal information. The results of this experiment show us the importance of correctly determining the SNR for pulse-like sounds and provide us with a point of reference in the form of maximum local SNR values for our current method for determining the SNR. This can be used in modeling human auditory perception and designing future experiments to study human perception of pulse-like and noise-like real-world sounds.

### Recognition Without Prior Knowledge

Although the differences between categories in detection and last correct recognition are highly significant, the differences in first correct recognition between the groups are not. As explained in the theoretical background, Ballas shows that many factors influence recognition of environmental sounds. Though the target sounds were selected to minimize differences due to these effects, these or other factors apparently still influenced the results of

the experiment and will have to be controlled in future experiments.

An interesting question concerning the effects influencing recognition of environmental sounds in noise is the following. Is there a difference in recognition threshold for sounds with more distinctive acoustical features and those which are more dependent on knowledge or inferences about properties of the source for correct recognition? To reason about physical properties of the sound source, enough information needs to be present in the noise-masked signal to recognize these properties. Therefore, we suspect that recognition thresholds for sounds with distinctive acoustical features are lower than those for sounds depending on inferences about source properties for correct recognition.

Research by Gygi et al. (2007) indicates that people do indeed listen to sounds in terms of physical properties of the sound source as well as acoustical properties of the sound. This is illustrated by the fact that when people are asked to rate the similarity of two sounds, sounds from similar sources are rated more similar than can be explained by the acoustical similarity. People seem to base their judgement both on properties of the sound itself and on knowledge of the sound source (Gygi, 2007).

We find further support for our intuition in the research of Ballas (1993). Ballas described the effects of what he called causal uncertainty; the uncertainty about the origin of the sound based on the number of sources that produce similar sounds. Sounds for which this causal uncertainty is high depend on inferences about source properties for correct recognition. High causal uncertainty results in low reaction times for recognition of the sounds. We argue this is due to the top-down processes involved when it is necessary for correct recognition to reason about source properties.

### Improvements to the experimental design

The results and the discussion above led to a number of improvements to the experimental design. For example, to obtain more informative results on first recognition it is necessary to minimize differences in recognizability of the target sounds used in the experiment. Familiarity with the sound and typicality compared to a mental stereotype are effects of great influence on the recognizability of a sound. A way to counter these effects would be to first determine familiarity and typicality of possible stimulus sounds, or alternatively, have a group of naive listeners score the sounds for recognizability and choose only those sounds with high familiarity and typicality, or a high recognizability score. The use of recognizable sounds will lower the variance in first correct recognition, by minimizing the chance of participants having difficulty naming the sounds while the acoustic evidence is sufficient for disambiguation.

Another improvement to the experiment is to ensure participants are presented with at least one, preferably two or more, sounds from each category. This will result in more useful data from an equal number of participants. Finally, pulse-like sounds need to be presented to the

participants starting from an even lower computed maximum local SNR, -20 or even -25dB.

## Conclusion

The results of this experiment show that the problem of determining the SNR for pulse like sounds is more important than we initially assumed. Evidently, the auditory system incorporates both temporal and spectral integration of evidence for the presence of sounds in noise. Current models for human auditory perception of noise-masked signals focus on temporal integration of evidence. This has provided a good model for the detection of tonal sounds, pitch perception and speech perception. However, environmental sounds include many pulse-like and noise-like sounds. A good model of environmental sound perception should therefore include spectral integration of information.

Though the realization that there is need for a more accurate method to determine the SNR of pulse-like sounds implies we cannot draw any further conclusions about the differences in human perception of pulse-like, noise-like and tonal sounds, this realization in itself is a valuable conclusion. The experiment provides valuable data for use in modeling efforts and further experiments to understand how humans integrate and use the available acoustic evidence.

## References

- Allen, J.B. (1994). How do humans process and recognize speech? *IEEE Transactions on Speech and Audio Processing* 2(4), 567--577.
- Andringa, T.C. (2002). *Continuity Preserving Signal Processing*. PhD dissertation, University of Groningen.
- Andringa, T.C., & Violanda, R. (2008). The texture of natural sounds. *In proceedings of acoustics '08*, (pp. 3141-3146). Paris, France.
- Ballas, J.A. (1993). Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19(2), 250-267.
- Gaver, W.W. (1993). What in the World Do We Hear?: An Ecological Approach to Auditory Event Perception. *Ecological Psychology*, 5(1), 1-29.
- Gygi, B., Kidd, G.R., & Watson, C.S. (2007). Similarity and categorization of environmental sounds. *Perception and Psychophysics*, 69(6) 839-855.
- Hawkins, J.E., & Stevens, S.S. (1950). The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America*, 22, 6-13
- Nábělek, A., & Robinson, P. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *Journal of the Acoustical Society of America*, 71, 1242-1248.