AU DITORY CUES DETERMINING THE PERCEPTION OF THE SIZE AND SPEED OF ROLLING BALLS

Mark M.J. Houben (1), Armin Kohlrausch (1,2), Dik J. Hermes (1)

(1) IPO, Center for User-System Interaction
P.O. Box 513, NL-5600 MB Eindhoven The Netherlands
(2) Philips Research Laboratories Eindhoven
Prof. Holstlaan 4, NL-5656 AA Eindhoven, The Netherlands
m.m.j.houben@tue.nl

ABSTRACT
This study investigates the auditory perception of the size and the speed of rolling balls. Prior experiments showed that subjects can discriminate differences in size and speed of wooden rolling balls on the basis of recorded sounds. Recorded sounds were manipulated by merging the temporal characteristics of one sound with the spectral characteristics of another. Perception experiments showed that when subjects had to choose the larger ball from two sounds, they had a preference for the spectral content of a large ball. If subjects had to choose the faster out of two sounds, they preferred the spectral content of a small ball, and, to a lesser degree, the spectral content of a fast rolling ball. The temporal cues in the sounds were of minor importance for the range of stimuli used in this experiment, possibly because sounds with much amplitude modulation and bouncing were excluded from the experiments.

1. INTRODUCTION
By listening to sounds in everyday life, people are able to extract information about the sound source, the location, and the environment in which the sound is produced. For example, people can hear from the sound of a car, whether it is big or small, approaching or leaving, far away or nearby, driving on a dry or wet street, etc. [1]. Although nonspeech sound is a familiar and natural medium to convey information, it is barely used in systems based on information technology.

In order to create suitable auditory interfaces we have to better understand how people perceive everyday sounds. We have chosen to study the sounds of rolling balls because the rolling ball can be used as a metaphor for cursor movement resulting from moving a mouse or a trackball. After presenting some prior experiments on the perception of the size and speed of a rolling ball, it is investigated what kind of auditory information within the sounds of rolling balls is used by naive subjects to judge their size and speed. In a later stage we intend to synthesize parameterized sounds of rolling balls whereby information to be displayed can be mapped to dimensions of the rolling ball, such as speed and size.

2. PRIOR EXPERIMENTS

2.1. Perception of size at constant speed and speed at constant size

In previous experiments subjects were asked to discriminate differences in the size of rolling wooden balls as well as differences in the speed of rolling on the basis of recorded sounds [2]. The results showed the ability of subjects to discriminate between the sounds of wooden rolling balls of different sizes. Subjects were generally also able to discriminate between the sounds of rolling balls with different speeds. However, some subjects had difficulties labeling the speed correctly, probably because no feedback about the correctness of the responses was provided. Those subjects were able to discriminate the speed of a rolling ball, but consistently mistook the slower ball for the faster one. Supplying feedback about the correctness of the response might solve this, but then we would not know whether the participant would indeed listen to the speed of the balls.

2.2. Interaction between size and speed

In a following experiment, the interaction between size and speed was tested [3]. Subjects listened to pairwise recordings of wooden balls rolling over a wooden surface. Both size and speed of the balls were varied over three levels. In one experiment subjects had to decide which of the two sounds in a pair was produced by the larger ball, whereas in another experiment the task was to decide which of the two sounds in a pair was produced by the faster rolling ball. The results showed that, for the size judgment task, the percentage correct responses, Pc, lay around 90%, which is above the upper boundary of the 95% prediction interval for guessing (Pc of 58%). Judging the speed was not as easy as judging the size of rolling balls, and it depended on the changes in size. An increase in speed accompanied by a decrease in size improved the identifiability of speed while subjects had more difficulties identifying the speed when both speed and size were increased. The results indicated that when both the size and speed of a rolling ball are varied, subjects generally are still able to discriminate the size and speed, but that the judgments are influenced by an interaction effect between the two physical properties of the balls.

3. AUDITORY CUES

The interaction effect encountered in the previous experiment, when both size and velocity of a rolling ball are increased, may be caused...
velocity experiments described in Section 2.1 are shown in Fig-

ure 1. The grey squares depict the values for the stimuli from the size experiment. The abscissa located on top of the figure gives the diameter of the ball for these stimuli. Values for the stimuli from the speed experiment are depicted by black circles and the abscissa at the bottom of the figure gives the velocity of the ball for these stimuli. The centroid is clearly influenced by size as well as speed, though the latter influence is smaller. Furthermore, the centroid increases with increasing speed and decreases with increasing size.

Investigation of auditory cues that subjects may use when judging the size and speed of rolling balls indicate that these cues are ambiguous. For example, an increase in the centroid of the specific loudness may be induced by a decrease in size or an increase in speed. However, the centroid is only a rough measure of spectral shape and, for example, does not take the spectral tilt into account. The influence of speed on the centroid of specific loudness is smaller than size which agrees with the findings of previous experiments, namely that for the chosen stimuli discrimination of size is easier than discrimination of speed. A convincing temporal cue was not found. However, the number of ticks per second present in the signal varied slightly with increasing speed.

4. PERCEPTION EXPERIMENTS WITH MANIPULATED SOUNDS

In previous studies [2, 3] it appeared that an interaction exists between the perception of size and speed of rolling balls. Small balls are confused with fast rolling balls. This interaction also emerged by analyzing the spectral and temporal cues subjects could use when judging the size and speed of rolling balls, such as discussed in the previous section. To obtain more information about the interaction effect, it was studied what kind of acoustic information subjects use when judging the size and speed. For this purpose recorded sounds were manipulated by combining the temporal characteristics of one sound with the spectral characteristics of another. These modified sounds served as a basis for perception experiments, which might help to unravel the perceptual cues for size and speed.

4.1. Sound-manipulation algorithm

The general approach consists of combining the spectral envelope of one sound, $s_1$, with the temporal envelope of another sound, $s_2$, as illustrated in Figure 2. First the two sounds were filtered with a Gammatone filterbank [7] with 32 channels regularly spaced on an ERB scale from 20 Hz to 24 kHz (half the sample frequency) resulting in $s_{1,c}$ and $s_{2,c}$ with channel index $c = 1 \cdots 32$. Per channel the Hilbert envelope was calculated, resulting in $e_{1,c}$ and $e_{2,c}$. The new signals per channel, $s_{12,c}$, were synthesized by substituting the temporal envelopes of signal one by the temporal envelopes of signal two (left fraction) but maintaining the spectral energy levels of signal one (right fraction):

$$s_{12,c} = s_{1,c} \cdot \frac{e_{2,c}}{e_{1,c}} \cdot \frac{\xi_{1,c}}{\xi_{2,c}}$$

with $\xi_{1,c}$ and $\xi_{2,c}$ the mean values of envelopes $e_{1,c}$ and $e_{2,c}$, respectively. The new sound was obtained by summing all signals in the channels and compensating for the group delay $\tau_c$ of the filters:

$$s_{12}(t) = \sum_{c=1}^{32} s_{12,c}(t + \tau_c).$$
in which $t$ denotes time. It is expected that $s_{12}$ combines the spectral characteristics of $s_1$ and the temporal characteristics of $s_2$.

### 4.2. Method

Sound recordings of wooden balls rolling over a wooden surface were used. Variation of size (diameters of 35 mm and 55 mm) and speed (approximately 0.60 m/s and 0.85 m/s), both at two levels, resulted in four sounds: small and slow, small and fast, large and slow, and large and fast. In this study, we did not select stimuli on the basis of their temporal content. As a result, the stimuli were not completely without amplitude modulation and bouncing. In the sound produced by the small ball with a diameter of 35 mm, irregular ticks can be heard which are not present in the sounds created by the larger ball with a diameter of 55 mm. In the sound produced by the large ball rolling slowly, some amplitude modulation can be heard which is not present in the other sounds.

New sounds were synthesized by combining the spectral content of one sound with the temporal content of another sound (see previous section). In this way 16 stimuli were obtained from the four original recordings\(^2\). The stimuli were presented pairwise, in random order, over headphones to 10 naive subjects seated in a soundproof booth. The duration of the stimulus was 800 ms with 700-ms silence in between. Only comparisons between different stimuli were made, resulting in 240 pairs of stimuli (the entire stimulus set minus the diagonal). The pairs were only played once and no feedback about the correctness of the responses was given.

The experiment was performed twice, with a difference in task: In one set, subjects had to decide which of the two sounds in a pair was produced by the largest ball. In another set, subjects had to decide which of the two sounds in a pair was produced by the fastest rolling ball.

Four parameters with two levels describe one single stimulus:
- **sizeSpec**: the size (small or large) of the rolling ball providing the spectral content,
- **veloSpec**: the velocity (slow or fast) of the rolling ball providing the spectral content,
- **sizeTemp**: the size (small or large) of the rolling ball providing the temporal content,
- **veloTemp**: the velocity (slow or fast) of the rolling ball providing the temporal content.

### 4.3. Results

The homogeneity of the responses among subjects was checked by constructing an intercorrelation matrix using Pearson coefficients. This matrix revealed that for the size judgment task all ten subjects correlated uniformly and significantly with one another with respect to their responses. The intercorrelation matrix for the speed judgment task revealed that the responses of one out of ten subjects did not correlate significantly at the 0.01 level with the results of any of the other subjects. The responses of this single subject was therefore not taken into account in the further analysis.

---

\(^2\)Sound examples s1.wav and s2.wav on the CD-ROM are original recordings of the small and slow ball, and large and fast ball, respectively. Sound examples s12.wav and s21.wav are synthesized by combining the spectral content of s1 with the temporal content of s2 and vice versa.
5. STATISTICAL ANALYSIS AND DISCUSSION

If subjects only attend to the spectral cues induced by size when judging the size of rolling balls, and they do this perfectly, the results would be 1 for sizeSpec, and 0.5 for all other parameters in Figure 3. It can be seen that the average results do not differ much from these values, revealing that when subjects had to choose the larger ball, they chose the sound with the spectral content of a large ball. The influence of the spectral content induced by speed (veloSpec) as well as the temporal content of the sound (sizeTemp and veloTemp) was much less. To search for statistically significant effects within the data, two procedures were followed. First, to get an overall impression of important effects, t-tests were conducted. Second, to test interaction effects, a binary logistic regression was applied to the data.

T-tests were conducted to determine if the data in Figure 3 differed from chance (a proportion of 0.5). The probability of obtaining the given result supposing that subjects merely guess, is given per parameter in the second row of Table 1. These values were compared to a significance level of 0.05/4 = 0.0125 (Bonferroni adjustment) to ensure that after doing all four tests, the chance of detecting at least one significant difference from guessing while there was no difference is limited to 5 percent. In the table, values below 0.0125 (implying a result significantly different from guessing) are italicized. Only sizeSpec is statistically significant at the Bonferroni adjusted 5 percent level, indicating that subjects were able to choose the larger ball by attending mainly to the spectral cues induced by size.

Figure 4, with the results of the speed judgment task, shows, again, that sizeSpec is most important. The proportion “faster” judgments of sizeSpec is close to 0, which indicates a preference for the spectral content of a small ball when subjects were asked to choose the faster ball. Additionally, subjects had a preference for the spectral content of a fast ball, as the proportion “faster” judgments for sizeSpec is above chance. Results of t-tests, shown in Table 1, confirmed that sizeSpec and veloSpec are significant at the Bonferroni adjusted 5 percent level.

To analyze the results of the experiments more closely, in particular interaction effects, a binary logistic regression was applied to the data. This type of regression was chosen because in our experiments the response variable was binary ( preference for first or second stimulus of a pair, 1 or 0) whereas in linear regression the response variable is assumed to be normally distributed and may lead to predictions of the response variable taking values other than 0 or 1. The response variable is transformed from the observed probability ($0 \leq p \leq 1$) into a variable with values between $-\infty$ and $+\infty$ by using the logit transformation of $p$:

$$\logit(p) = \ln \frac{p}{1-p},$$

(4) which is the natural logarithm of the odds ratio [8]. Binary logistic
regression models the logit transformation of the observed probability as a linear function of the explanatory variables. In our experiment we have four parameters describing the stimuli, namely sizeSpec, veloSpec, sizeTemp, and veloTemp, for simplicity denoted as par₁ to par₄. These parameters can take the values -1 (‘low’, i.e. small or slow) and 1 (‘high’, i.e. large or fast). A basic model for a stimulus characterized by parameters par₁ to par₄ is assumed to be of form

\[
\logit(p) = \beta_1 \cdot par_1 + \beta_2 \cdot par_2 + \beta_3 \cdot par_3 + \beta_4 \cdot par_4 + \beta_12 \cdot par_{12} + \beta_{13} \cdot par_{13} + \ldots + \beta_{1234} \cdot par_{1234} \tag{5}
\]

with \(par_{12} = par_1 \cdot par_2, par_{13} = par_1 \cdot par_3, par_{123} = par_1 \cdot par_2 \cdot par_3, \) etc. Because the stimuli were presented pairwise, a set of explanatory variables for main effects was constructed by taking the difference in parameter values between the first (\(par_{1,x}\) to \(par_{4,x}\)) and second stimulus (\(par_{1,x} to par_{4,x}\)) resulting in \(dpar_{1} to dpar_{4}\) with values -2, 0, or 2. In the same way, explanatory variables for interaction effects were constructed by taking the difference in multiplied parameter values between the first and second stimulus. The complete model can now be written as

\[
\logit(p_{i,t}) = \beta_1 \cdot dpar_{1} + \beta_2 \cdot dpar_{2} + \beta_3 \cdot dpar_{3} + \beta_4 \cdot dpar_{4} + \beta_{12} \cdot dpar_{12} + \beta_{13} \cdot dpar_{13} + \ldots + \beta_{1234} \cdot dpar_{1234} \tag{6}
\]

with \(p_{i,t}\) the proportion of subjects that chose the first stimulus when presenting the pair consisting of stimulus \(s\) and stimulus \(t\), and \(dpar_{1} = par_{1,x} - par_{1,t}, dpar_{2} = par_{12,x} - par_{12,t} = par_{1,x} \cdot par_{2,x} - par_{1,t} \cdot par_{2,t}, \) etc. A backward stepwise selection method based on the likelihood ratio statistic was used for reducing this model. As initial model the complete model with all explanatory variables for main effects was constructed by taking the difference in multiplied parameter values between the first and second stimulus. The most important parameter appears to be sizeSpec, indicating that subjects mainly chose the sound with the spectral content of a large ball when they had to choose the larger ball. Furthermore, sizeTemp and the interaction between sizeSpec and sizeTemp slightly influence the size judgments of the subjects. Substituting the estimated coefficients into the model for individual stimuli (Equation 5) reveals a measure of perceived size (ranging from \(-\infty\) or infinitesimal small to \(+\infty\) or infinitesimal large):

\[
\text{perceived size} = 0.988 \cdot \text{sizeSpec} - 0.233 \cdot \text{sizeTemp} - 0.252 \cdot \text{sizeSpec} \cdot \text{sizeTemp} \tag{7}
\]

Figure 5 visualizes the perceived size for the size judgment task as a function of sizeSpec and sizeTemp. It shows that if the ball providing the spectral content is large (sizeSpec = large), the sound is judged as being produced by a large ball, independently of the value of sizeTemp. On the other hand, if the ball providing the spectral content is small (sizeSpec = small), the sound is judged as being produced by a small ball and this percept is even stronger if the sound also contains the temporal content of a small ball (sizeTemp = small).

Apparently, the judgment of size depends primarily on parameters arising from size (sizeSpec and sizeTemp) and is less influenced by the speed of the rolling ball, which agrees with the results of the size judgment task of the interaction experiment described in Section 2.2. Since Figure 5 shows that the perception of size is dominated by the spectral attributes of size (sizeSpec), and the variation in spectral cues induced by size is much larger than by speed (see Figure 1 in Section 3), it is no surprise that variations in speed hardly influence discrimination of size.

The binary logistic regression applied to the pooled results of the speed judgment task leads to a model with four terms: difference in sizeSpec (with an estimated value of the corresponding coefficient \(\beta = 0.988\)), difference in sizeTemp (\(\beta = 0.235\)), and the interaction between these two (\(\beta = 0.252\)) when the responses of the subjects were not pooled but binary logistic regressions were applied to the individual responses instead, exactly these three came out as significant for all subjects. This model with three terms for subjects were not pooled but binary logistic regressions were applied to the individual responses instead, exactly these three came out as significant for all subjects. This model with three terms for interactions was constructed by taking the difference in multiplied parameter values between the first and second stimulus. The overall percentage of correctly predicted individual responses was 73%.

The most important parameter appears to be sizeSpec, indicating that subjects mainly chose the sound with the spectral content of a large ball when they had to choose the larger ball. Furthermore, sizeTemp and the interaction between sizeSpec and sizeTemp slightly influence the size judgments of the subjects. Substituting the estimated coefficients into the model for individual stimuli (Equation 5) reveals a measure of perceived size (ranging from \(-\infty\) or infinitesimal small to \(+\infty\) or infinitesimal large):

\[
\text{perceived speed} = -0.812 \cdot \text{sizeSpec} + 0.536 \cdot \text{veloSpec} + 0.292 \cdot \text{sizeSpec} \cdot \text{veloSpec} + 0.189 \cdot \text{sizeTemp} \tag{8}
\]

Figure 6 visualizes the perceived speed for the speed judgment
task as a function of sizeSpec and veloSpec. It shows that if the ball providing the spectral content is large (sizeSpec = large), the sound is judged as being produced by a ball rolling slowly. On the other hand, if the ball providing the spectral content is small and fast (sizeSpec = small, veloSpec = fast), the sound is judged as being produced by a fast rolling ball, whereas, if the ball providing the spectral content is small and slow (sizeSpec = small, veloSpec = slow), the sound is perceived as being neither slow or fast. The small difference in perceived speed between a small ball rolling slowly (sizeSpec = small, veloSpec = slow) and a large ball rolling fast (sizeSpec = large, veloSpec = fast) indicates the difficulty shown by subjects to discriminate the speed between these two rolling balls. This agrees with the results of the speed judgment task of the interaction experiment described in Section 2.2. The fact that the variation of perceived speed is larger for size-Spec than for veloSpec corresponds to the higher influence of size on spectral cues compared to speed, as shown in Figure 1 in Section 3. Therefore, judgment of speed is much more difficult if size is varied simultaneously.

The final and smallest significant effect on judgment of speed, found by the binary logistic regression, was that of sizeTemp \((\beta = 0.189)\). The small positive coefficient indicates that if the sound contains the temporal content of a large ball, the ball is judged as being slightly faster than if the sound contains the temporal content of a small ball. The small effect of sizeTemp and the non-significance of veloTemp reveal that, for the range of stimuli used in this experiment, temporal aspects are of relatively minor importance to the listeners. However, the large interquartile range for sizeTemp points to a large difference between subjects. Possibly some subjects do take temporal cues into account.

In this study, we deliberately did not select stimuli on the basis of their temporal content. In future experiments, we plan to add amplitude modulation to the sounds to analyze the influence on the perception of size and speed and to address the question whether subjects attend to the linear or angular velocity of a rolling ball. The results will serve as a basis for the synthesis of rolling sounds, for which certain parameters can be adjusted in such a way that the listener can perceive the sounds as those of balls of well-defined sizes, rolling with well-defined speeds.

6. CONCLUSIONS

The results of this paper and previous experiments lead to the following description of auditory judgments of size and speed of rolling balls.

- The judgment of size as well as speed is dominated by spectral cues for stimuli without much amplitude modulation and bouncing.
- The variation in spectral cues induced by size is much larger than by speed.
- The ability of judging the size is hardly affected by independently varying speed.
- Judgment of speed is much more difficult if size is varied simultaneously.

8. REFERENCES