AUDITORY 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][1]. 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

[1] Sound examples size1.wav and size2.wav on the CD-ROM are produced by two wooden balls of different sizes (diameter of 55 mm and 68 mm, respectively) rolling at the same speed (0.80 m/s). Sound examples speed1.wav and speed2.wav are produced by a ball with a diameter of 45 mm rolling at two different speeds (0.71 m/s and 0.87 m/s, respectively).

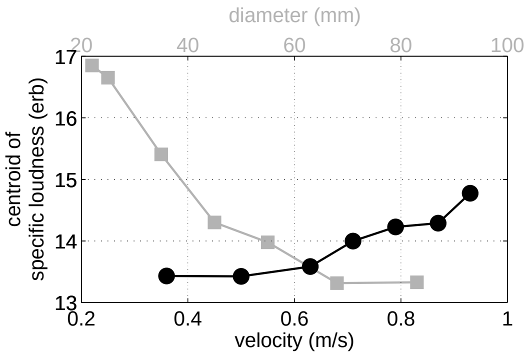

Figure 1: *Centroid of specific loudness for the stimuli used in the size perception experiment (gray squares) and speed perception experiment (black circles).*

by the fact that changes in these two physical parameters affect auditory cues which are used by the subjects, in a similar way.

Two possible temporal cues are amplitude modulation induced by the ball's imperfect sphericity, and a kind of rough ticks, irregularly distributed in time, which are probably induced by the bouncing of the ball or by irregularities on the surface of the rolling ball or the supporting surface. The recorded sounds used in the experiments [3] were chosen on the basis of the absence of distinct amplitude modulation and bouncing. Therefore only a very weak positive correlation between the number of ticks and the speed could be observed for these sounds.

Both variations in size and speed of the rolling ball induce changes in spectral shape of the sound, including changes in the proportion of low and high frequencies, and changes in spectral tilt, which is the overall slope of the spectrum. As a first step, the spectral content of a sound is expressed in one figure by calculating the centroid of the specific loudness. Specific loudness, which is the loudness per critical band, takes the frequency specific absolute threshold of our ears into account, and is a kind of 'loudness density'. The calculation is based on a model for computing loudness by Moore *et al.* [4]. From this the centroid of specific loudness is calculated by:

$$\frac{\int \theta \; N'(\theta) \; d\theta}{\int N'(\theta) \; d\theta}, \tag{1}$$

in which N' is the specific loudness, and θ is the ERB (Equivalent Rectangular Bandwidth) rate. The denominator, the integral over the specific loudness contour, is an estimate of the overall loudness of a given sound in sone.

Equation 1 can be seen as a combination of the spectral centroid, which is the center of gravity of the power spectrum of the sound and which is related to the perceived 'brightness' [5], and the definition of 'sharpness' proposed by Zwicker and Fastl [6]. The spectral centroid uses a linear frequency axis instead of the logarithmic ERB-rate scale and it does not take the frequency specific absolute threshold of the human ear into account. Sharpness uses a critical-band-rate scale (unit 'Bark') instead of an ERB-rate scale and includes an extra weighting factor that takes into account that sharpness of narrow-band noises increases unexpectedly strongly at higher frequencies.

The results of this analysis for the stimuli used in the size and velocity experiments described in Section 2.1 are shown in Figure 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):

$$s_{12,c} = s_{1,c} \cdot \frac{e_{2,c}}{e_{1,c}} \cdot \frac{\overline{e}_{1,c}}{\overline{e}_{2,c}}, \tag{2}$$

with $\overline{e}_{1,c}$ and $\overline{e}_{2,c}$ 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 τ_c of the filters:

$$s_{12}(t) = \sum_{c=1}^{32} s_{12,c}(t + \tau_c), \tag{3}$$

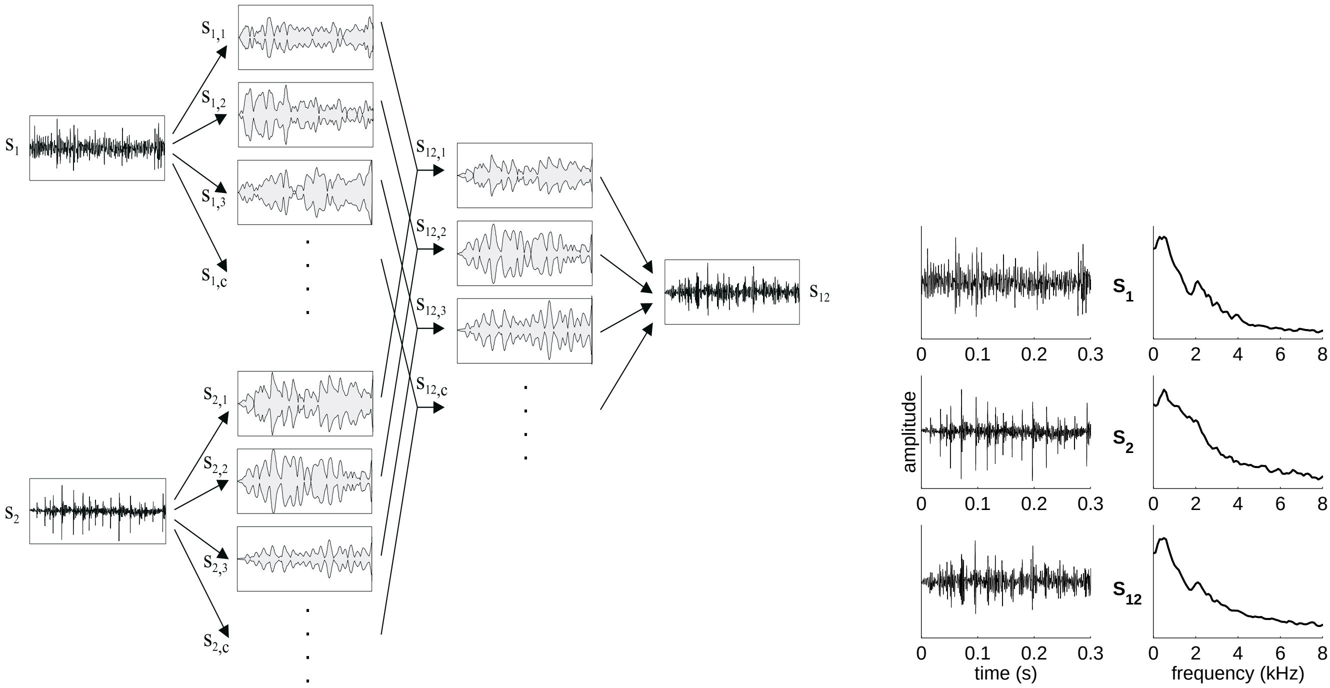

Figure 2: *The left panel shows the sound-manipulation algorithm. The signals are plotted in the time domain.* s_1 *and* s_2 *are two real recordings.* $s_{1,c}$ *and* $s_{2,c}$ *are the Gammatone-filtered signals in the channel with index* c *for signal* s_1 *and* s_2*, respectively.* $s_{12,c}$ *are the new signals in the* c*th channel, synthesized by substituting the temporal envelopes of* $s_{1,c}$ *by the temporal envelopes of* $s_{2,c}$*.* s_{12} *is obtained by summing* $s_{12,c}$ *over all channels, and combines the spectral characteristics of* s_1 *and the temporal characteristics of* s_2*. The right panel shows signal* s_1*,* s_2 *and the resulting signal* s_{12}*, in both time and frequency domain.*

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 sounds 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 stimuli 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.

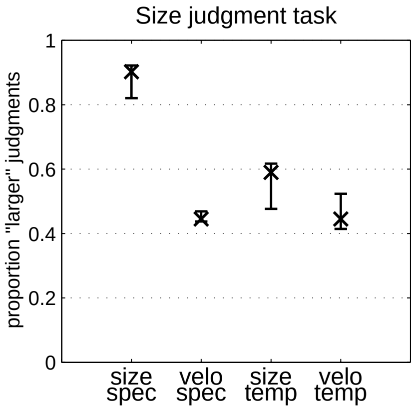

Figure 3: *Results of the size judgment task. For each parameter (denoted by the labels on the x-axis) the pairs with a difference in that parameter are considered and the proportion that subjects chose the higher value is visualized. Medians (crosses) as well as interquartile ranges (vertical bars) are shown.*

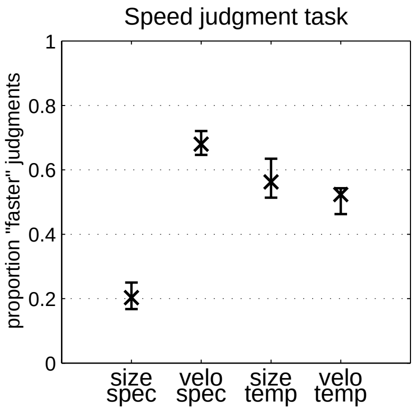

Figure 4: *Results of the speed judgment task. See Figure 3 for a description of the content.*

Figures 3 and 4 visualize the results of the size judgment task and speed judgment task, respectively. Per parameter only the pairs with a difference in that parameter are considered, and the proportion of choosing the stimulus with the higher value is shown. The x-axis labels denote the parameters that are considered. From left to right: difference in sizeSpec (i.e. stimulus pairs with a difference in *size* of the rolling ball which provides the *spectral* content), veloSpec, sizeTemp, and veloTemp. The y-axis denotes the number of times (in proportion to the total pairs examined) that subjects chose the higher value (larger or faster ball) of the parameter considered. Medians of subjects are shown by black crosses. Vertical bars denote the interquartile range. A proportion below 0.5 simply means that subjects preferred the lower value (small or slow) of the parameters considered. The larger the deviation from 0.5 (no preference), that is, the closer to 0 or 1, the stronger the preference for the lower or higher value, respectively.

Table 1: *T-test probabilities of obtaining the data in Figures 3 and 4 (size and speed judgment task, respectively) supposing that subjects merely guess.*

	sizeSpec	veloSpec	sizeTemp	veloTemp
size	*.000*	.017	.046	.11
speed	*.000*	*.000*	.088	.53

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:

$$\mathrm{logit}(p) = \ln \frac{p}{1-p}, \qquad (4)$$

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 par_1 to par_4. 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 par_1 to par_4 is assumed to be of form

$$\begin{aligned}\mathrm{logit}(p) = \beta_1 * par_1 + \beta_2 * par_2 + \beta_3 * par_3 + \beta_4 * par_4 \\ +\beta_{12} * par_{12} + \beta_{13} * par_{13} + \ldots \\ +\beta_{123} * par_{123} + \ldots + \beta_{1234} * par_{1234} \end{aligned} \quad (5)$$

with $par_{12} = par_1 * par_2$, $par_{123} = par_1 * par_2 * par_3$, 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 ($par_{1,s}$ to $par_{4,s}$) and second stimulus ($par_{1,t}$ to $par_{4,t}$) 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

$$\begin{aligned}\mathrm{logit}(p_{st}) = \beta_1 * dpar_1 + \beta_2 * dpar_2 + \beta_3 * dpar_3 \\ +\beta_4 * dpar_4 + \beta_{12} * dpar_{12} + \beta_{13} * dpar_{13} + \ldots \\ +\beta_{123} * dpar_{123} + \ldots + \beta_{1234} * dpar_{1234} \end{aligned} \quad (6)$$

with p_{st} the proportion of subjects that chose the first stimulus when presenting the pair consisting of stimulus s and stimulus t, and $dpar_1 = par_{1,s} - par_{1,t}$, $dpar_{12} = par_{12,s} - par_{12,t} = par_{1,s} * par_{2,s} - par_{1,t} * par_{2,t}$, etc. A backward stepwise selection method based on the likelihood ratio was used for reducing this model. As initial model the complete model with all explanatory variables was taken and variables were removed iteratively if the probability of the likelihood-ratio statistic, which is based on the maximum partial likelihood estimates, was greater than 0.01.

The regression results for the size judgment task revealed many significant terms, with three terms clearly standing out: difference in sizeSpec (with an estimated value of the corresponding coefficient β of 0.988), difference in sizeTemp (β = 0.235), and the interaction between these two (β = -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 the size judgment task predicted proportion "larger" judgments of 0.88, 0.50, 0.62, and 0.50 for sizeSpec, veloSpec, sizeTemp, and veloTemp, respectively, which match the observed mean values of 0.87, 0.45, 0.58, and 0.46 (shown in Figure 3) fairly well. 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 -∞ or infinitesimal small to +∞ or infinitesimal large):

$$\begin{aligned}\text{perceived size} = 0.988 \text{ sizeSpec} + 0.235 \text{ sizeTemp} \\ -0.252 \text{ sizeSpec} * \text{sizeTemp}\end{aligned} \quad (7)$$

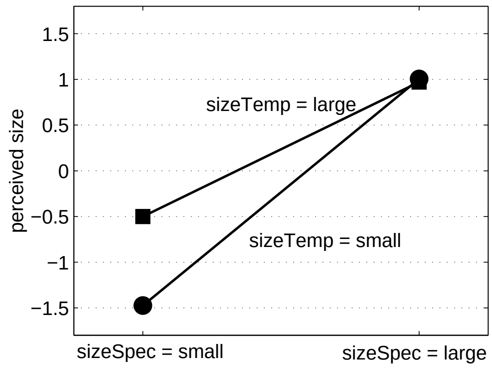

Figure 5: *Perceived size as function of sizeSpec and sizeTemp for the size judgment task.*

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 β of -0.812), difference in veloSpec (β = 0.536), the interaction between these two (β = 0.292), and difference in sizeTemp (β = 0.189). This model for the speed judgment task predicted proportion "faster" judgments of 0.16, 0.75, 0.50, and 0.50 for sizeSpec, veloSpec, sizeTemp, and veloTemp, respectively, which match the observed mean values of 0.20, 0.69, 0.57, and 0.51 (shown in Figure 4) fairly well. The overall percentage of correctly predicted individual responses was 72%.

Again, the most important parameter appears to be sizeSpec, but now with a negative coefficient indicating that subjects mainly chose the sound with the *spectral* content of a *small* ball when they had to choose the faster ball. Furthermore, veloSpec and the interaction between sizeSpec and veloSpec slightly influence the speed judgments of the subjects. Substituting the estimated coefficients into the model for individual stimuli (Equation 5) reveals a measure of perceived speed (ranging from -∞ or infinitesimal slow to +∞ or infinitesimal fast):

$$\begin{aligned}\text{perceived speed} = -0.812 \text{ sizeSpec} + 0.536 \text{ veloSpec} \\ + 0.292 \text{ sizeSpec} * \text{veloSpec} + 0.189 \text{ sizeTemp}.\end{aligned} \quad (8)$$

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 sizeSpec 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.

7. ACKNOWLEDGMENT

We would like to thank Ralph van Dinther and Preeti Rao for the implementation of the specific loudness model in the numerical software package MATLAB.

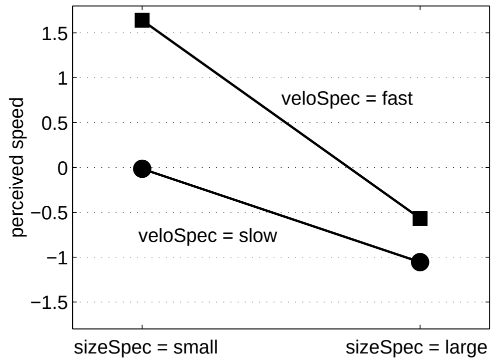

Figure 6: *Perceived speed as function of sizeSpec and veloSpec for the speed judgment task.*

8. REFERENCES

[1] W.W. Gaver, ”What in the world do we hear? An ecological approach to auditory event perception,” *Ecological Psychology*, vol. 5, pp. 1–29, 1993.

[2] M. Houben, L. Franssen, A. Kohlrausch, D. Hermes, and B. Eggen, “Auditory perception of the size and velocity of rolling balls,” in *Collected Papers from the joint ASA/EAA Meeting* (CDROM), Berlin, Germany, March 1999.

[3] M.M.J. Houben, D.J. Hermes, and A. Kohlrausch, “Auditory perception of the size and velocity of rolling balls,” in *IPO Annual Progress Report 34*, pp. 86–93, 1999.

[4] B.C.J. Moore, B.R. Glasberg, and T. Baer, “A model for the prediction of thresholds, loudness, and partial loudness,” *J. Acoust. Soc. Am.*, vol. 45, pp. 224–240, 1997.

[5] S. McAdams, J. Beauchamp, and S. Menneguzzi, “Discrimination of musical instrument sounds resynthesized with simplified spectrotemporal parameters,” *J. Acoust. Soc. Am.*, vol. 105, pp. 882–897, 1999.

[6] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, 2nd edition, Springer-Verlag, Berlin, Germany, 1999.

[7] R.D. Patterson, M.H. Allerhand, and C. Giguère, “Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,” *J. Acoust. Soc. Am.*, vol. 98, pp. 1890–1894, 1995.

[8] D.R. Cox, and E.J. Snell, *Analysis of binary data*, Chapman and Hall, London, UK, 1989.