**Theme of the Experiment:**

How is high level information about auditory sound organized?
How do humans categorize these sounds?

Studies of sound organization often find that a combination of causal, semantic and timbral factors contribute towards this task.

One of the most cited study in reference to this task is the Gaver Hypothesis's for categorizing sounds. He differentiates between 2 types of sound events, basic & complex. A basic level sound events are placed within each of the three physical states namely, solids, liquids and gases. Any sound that is produced due to interaction of more than one state of mass, are "hybrid" and are complex events. The hypothesis being that we categorize environmental sounds on the basis of what caused the sound rather than the sound itself.

In our research we have tried to explore an acoustic basis for this hypothesis, i.e can we find acoustic features which would perform a consistent classification of sounds as proposed.

**Experimental Setup (Based only on Gaver-3 data):**

Listeners were instructed to judge the event that caused each sound they heard. An example was given: “water dripping into a puddle” would be “dripping”. Listeners were encouraged to entertain more than one possible interpretation of a sound.

They were asked to explicitly judge the causality of the sounds by asking them to rate how likely it was that the sound was generated by a specified action. After hearing a single sound they were supposed to answer a series of questions about it, e.g., “How likely is it that this event was produced by dripping?” and so on for a total of 18 questions about the following actions: dripping, water sloshing around, splashing, an object being deformed, hitting, scraping, crushing, crumpling, breaking, bouncing, sawing, rubbing, filing, rolling, a puff of air or a gust, continuous airflow such as wind, pouring, an explosion.

All together this resulted in 18 questions presented in a row, in random order, for each sound. Subsequently a new sound, randomly chosen from the entire set of sounds was selected for the rating task. Ratings were entered by computer keyboard, on a scale of 1 (extremely unlikely) to 7 (extremely likely), with the option of re-hearing the current sound by hitting a designated key. Listeners received no feedback during the 75-minute test session.

In total there were 69 sounds artificially produced in the labs.

This experiment enabled representation of a sound as a 18 element array. The psychological distance between sounds, as measured by the distance between these arrays, was used to construct a similarity matrix within a cluster analysis, in order to determine whether the perceptual ratings of the sounds would produce interpretable clusters of the sounds. A hierarchical cluster analysis was used with “between-groups linkage”1 and squared-Euclidean distance (Gygi, Kidd & Watson 2007). There were six highly-distinguishable groups in the cluster analysis as reported:

1. Deformation
2. Explosion
3. Impact
4. Drip Pour
5. Wind- Roll-Rain
6. Scrape
(A more detailed scheme on actions associated with each sound can be found on page 18 of Lauries confidential draft)

**Acoustic Experiment:**
With help of labeled data obtained after cluster analysis of the psychological experiment, a 1 gaussian model was trained for classifying sounds into one of these categories.

MFCC's are extracted for acoustic representation of sounds. They are calculated for 25ms window of the signal, shifted by 10 ms each. 39 mel coefficients, and their first, second order differences are calculated to generate a 117D representation for each window. Thus each sound is represented as a sequence of 'n' 117D vectors (where n depends on number of windows).

Now with help of this data, mean and covariance are estimated for each of the 6 classes. Thus, we are able to model each class as 1 gaussian. Next, simple bayesian classification rule is used for categorizing each sound.

**Jack-Knife Experiment:**
1. Test set is constructed by randomly choosing 10 sounds (15% of available data). It is ensured that there is at least one sound from each class.
2. Gaussian model is trained on the remaining sounds and is used for classification on the test set.
3. This is repeated 50 times. (to reduce the noise in the classification)
4. The confusion matrix reported is a result of one of such experiment.

**Experiment on rest of the data:**
1. A model is trained on these 69 sound files and used for classifying rest of Laurie’s sounds.

From the results it seems that the following 3 classes cluster into one category:
1. 'D': Deformation
2. 'S': Scrape
3. 'P': Drip Pour