Low-Power Audio Classification For Hearing-Aids

David V. Anderson

Assistant Professor
Georgia Institute of Technology
dva@ece.gatech.edu

SPRP497

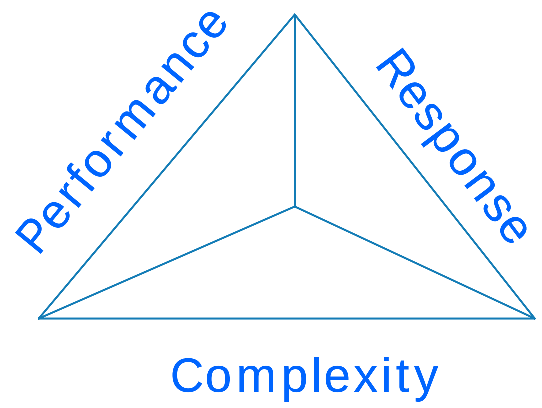
Project Objective

- **Automatic tuning of electro-acoustic response of hearing-aids to suit the audio environment**
  - We perform classification of the auditory environment to enable tuning of the hearing-aid
- **Requirements**
  - Small Size
  - Low-power
  - High accuracy
  - Low false alarms

TI Developer Conference

Desirable Qualities

- **Inter-class variability**
  - Features should provide good inter-class discrimination but still maintain intra-class cohesion
- **Features must be robust to noise**



- **Granularity Issue**
  - Trade-off between complexity of system and granularity of classes
- **Real-time response**
  - Computationally efficient classification structures and feature extraction algorithms

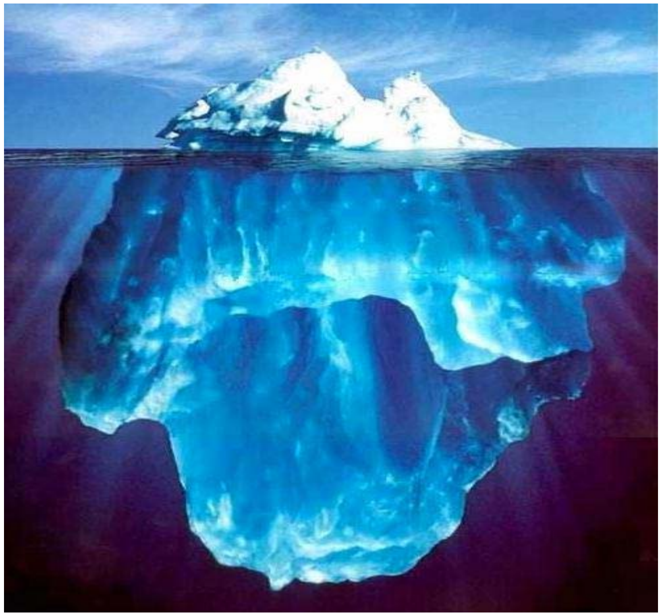
Some Current Classification Approaches

Methods	Accuracy	Complexity	Comments
GMM	Good	Moderate	Does not handle high dimensional data well
HMM	Good	High	Computationally expensive. Usually use GMMs for probability estimates.
SVM	Very Good	High	Computationally expensive. Is essentially a binary classifier.
Heuristics	Fair	Low	Easy to implement but accuracy in adverse conditions may not be very good.

Proposed Approach

- **Robust feature extraction**
  - Based on an advanced model of the human auditory system.
- **Very efficient algorithm for classification based on AdaBoost**
  - Final classifier can be implemented using MAC and a comparator.

Problems with Conventional Features

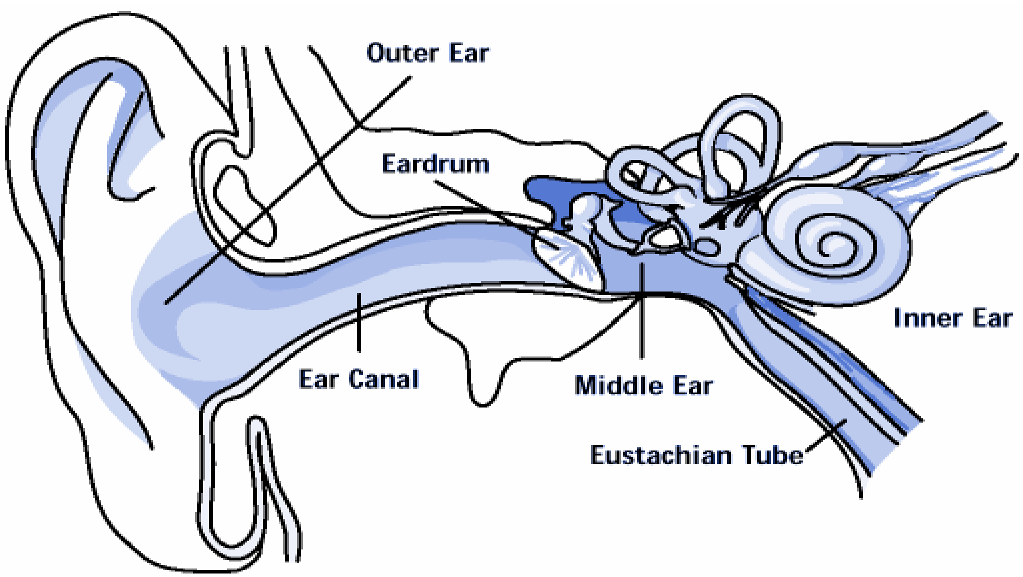


- **Work well in noise free case but performance degrades in presence of noise**
- **Accuracy is reduced greatly when different classes are presented simultaneously**

Why auditory modeling?

- **Humans do an extremely good job of classifying sounds**
- **Physiologically inspired perceptual features are**
  - Highly discriminative
  - Robust to noise

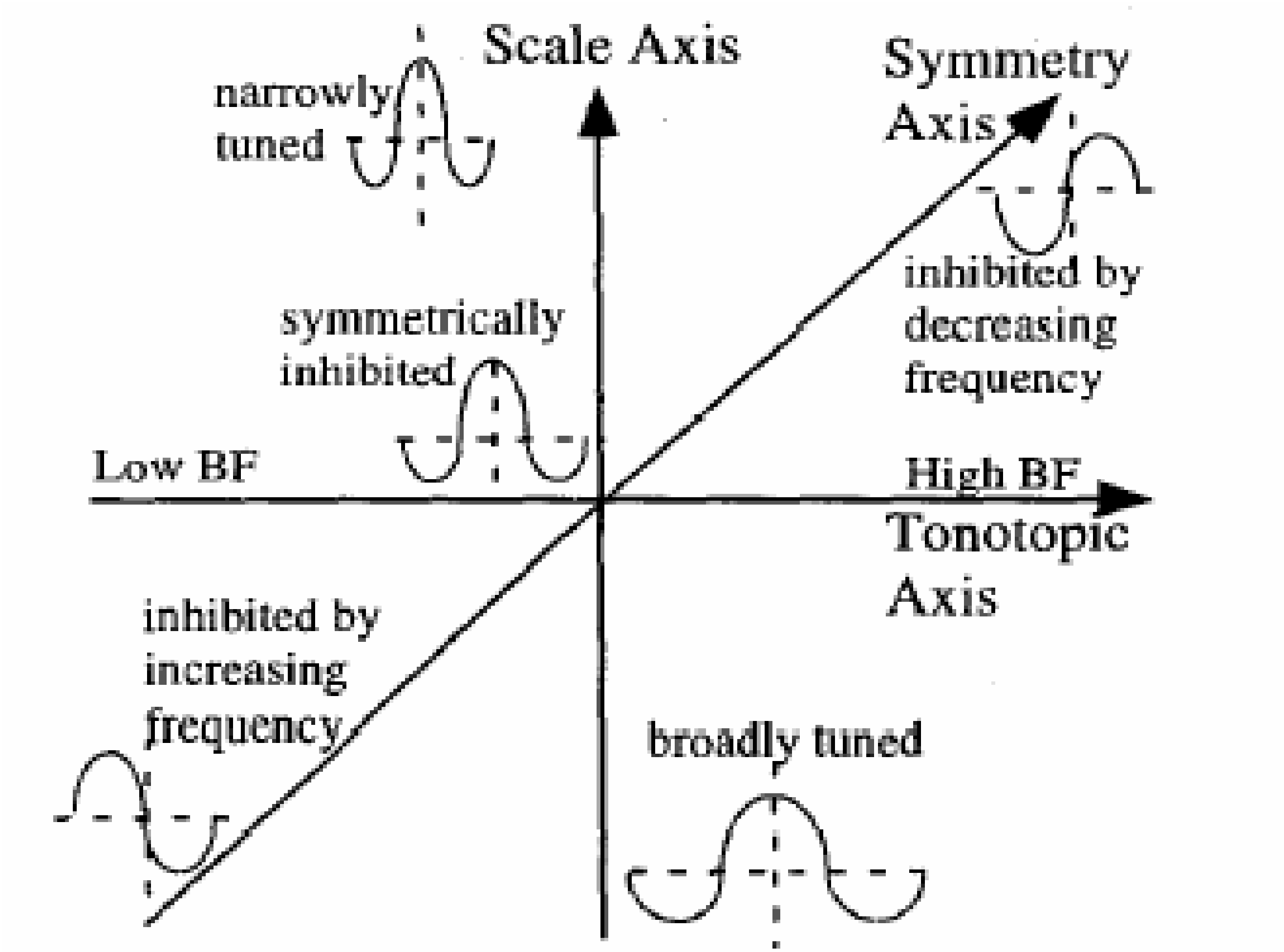
Auditory Modeling Based on Modulation Spectrum Theory



Input → $h(t;s)$ → ∂t → g( ) → w(t) → ∂s → v(s) → HWR → $\int_T$ →

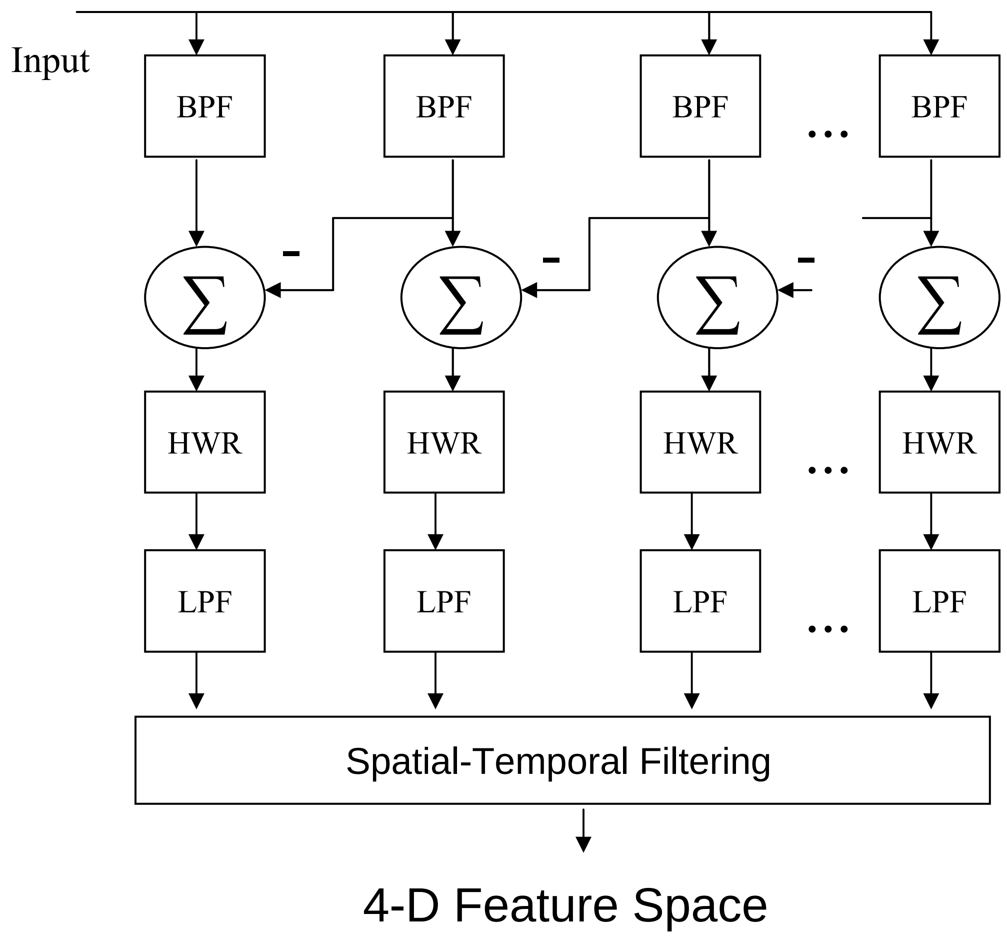
Cochlea ↔ Hair cell stage ↔ Cochlear nucleus

Response Fields of Neurons in the Cortex*



*Shamma et al.

Hardware Implementation



AdaBoost Classifier

- Given examples $(x_1, y_1),\ldots,(x_2, y_2)$ where $y_i = 0,1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = 1/(2m), 1/(2l)$ for $y_i = 0,1$ respectively, where m and l are the number of negatives and positives respectively.
- For t = 1 to T

1. Normalize weights,

$$w = w_{t,i} / (\Sigma_j w_{t,j})$$

2. Train h_j ; error, $\varepsilon_t = \sum_i | h_j(x_i) - y_i |$
3. Choose classifier h_t, with the least ε_t
4. Update weights:

$$w_{t+1,i} = w_{t,i} (\beta_t)^{(1-e_i)}$$

$e_i = 0$ if x_i if classified correctly, 1 otherwise

$$\beta_t = \varepsilon_t / (1 - \varepsilon_t)$$

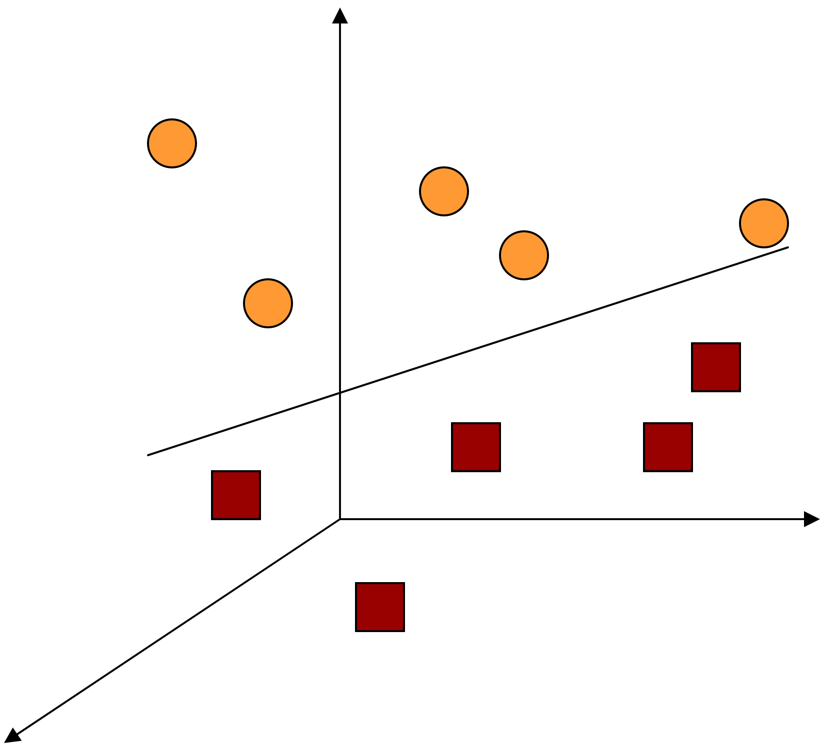
- The final strong classifier is:

$$h(x) = 1 \qquad \sum_{t=1}^{T} \alpha_t h_t(x) \geq (1/2) \sum_{t=1}^{T} \alpha_t$$

if

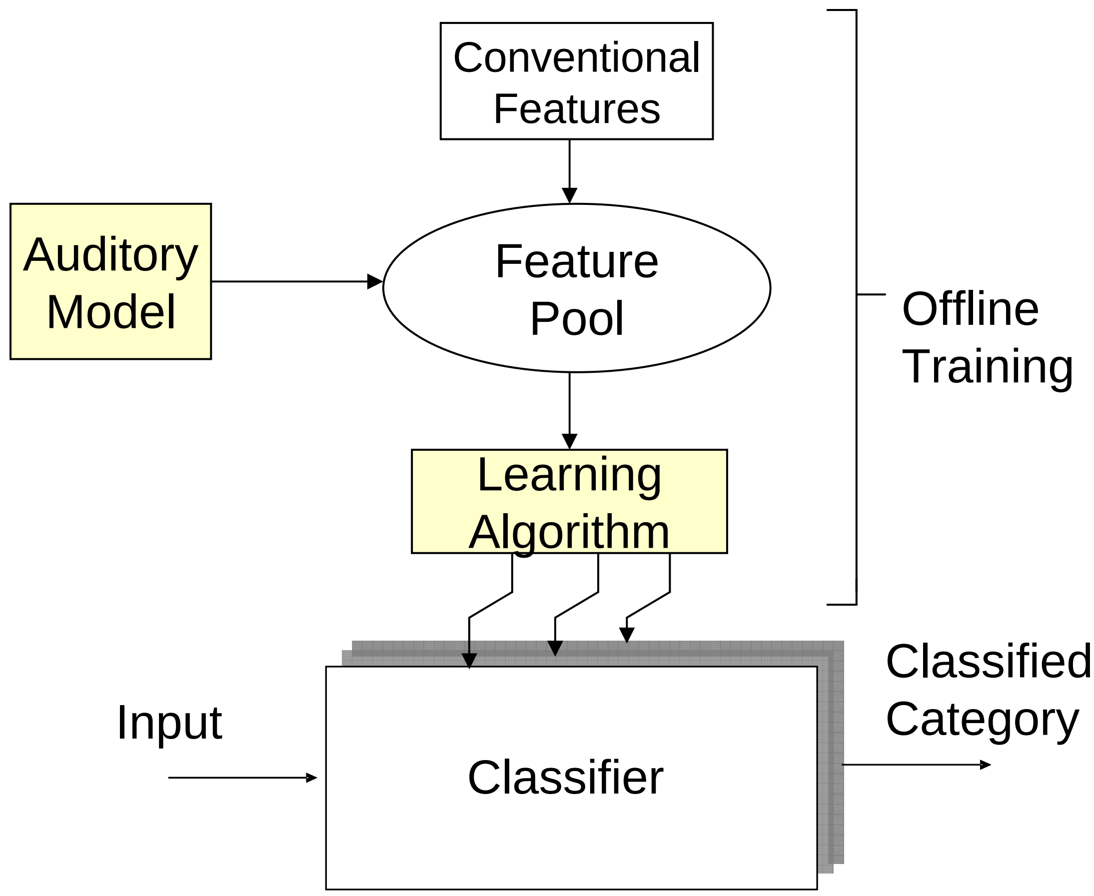
$= 0 \qquad$ else

where, α_t = log(1/β_t)



The AdaBoost classifier tries to find the decision boundary for the classification task by combining the multiple hypothesis based on single features.

Overall Structure of Proposed System

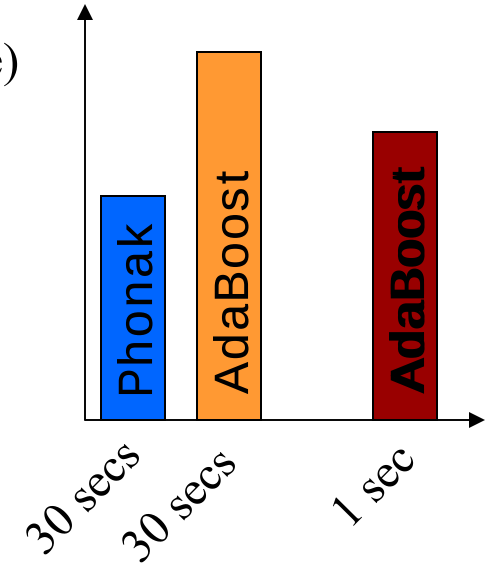


During offline training, the weights (alpha's) needed to combine the features to form the decision function are learned. The multi-class problem is structured as a combination of binary classification problems and the results are combined by majority voting.

Results - Matlab

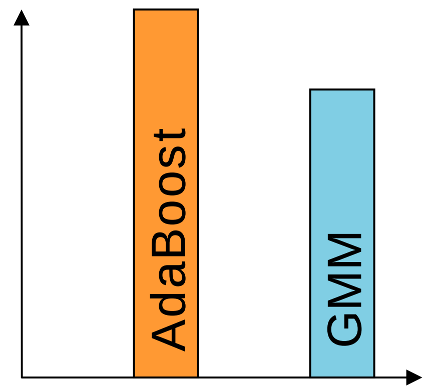
Phonak Database (Music, Speech, Noise, Speech in Noise)

	Phonak (30 sec data)	Version 1 (1 sec data)	Version 2 (30 sec data)
Overall	78.85 %	87.7 %	95.8 %



Tel-03 Database (Animal Vocalizations, Speech, Music, Noise)

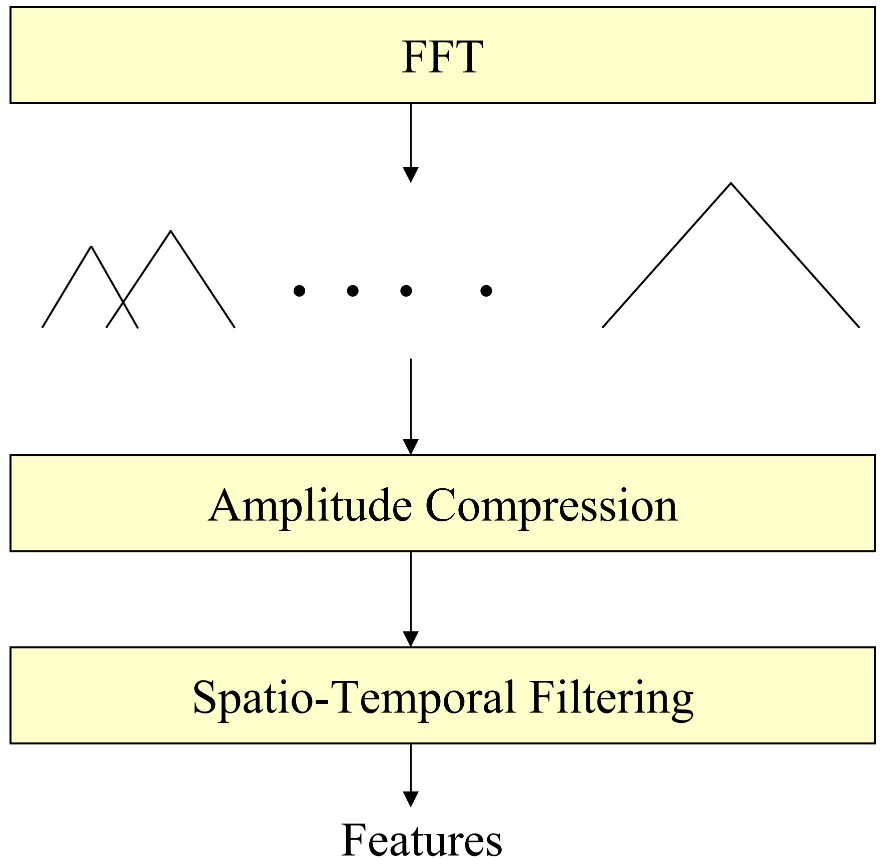
	GMM	AdaBoost
Overall	92.7 %	95.5 %



Hardware Implementation

- **In order to reduce the complexity of the feature extraction and to enable ease of implementation, some modifications were incorporated**
  - The frontend bandpass filters were replaced by an FFT and a mel-cepstra like processing was implemented to extract the auditory spectrum.

DSP Implementation



- **Simulation results with the new feature set showed no *overall* degradation in classification accuracy compared to the original feature set.**

Category	Original features	New Features (after modification for hardware implementation)
Music	**97.84 %**	**99.71 %**
Noise	**77.03 %**	**79.02 %**
Speech	**79.74 %**	**90.52 %**
Noisy Speech	**88.51 %**	**78.02 %**
Overall	85.96 %	86.82 %

Note: Drop in noisy speech performance is due to the use of FFT and mel-scale grouping.

C5510 Specifications

- **Sampling rate: 8 kHz**
- **For feature extraction (for 1 second segment):**
  - Size of data: 13 k words
  - Size of our code: 4 k words
  - Size of entire code: 16 k words
  - MIPS ?
- **For Classification**
  - 85 coefficients
  - 85 MACs, 85 additions and 1 compare

Low-Power Audio Classification For Hearing-Aids

David V. Anderson

Assistant Professor
Georgia Tech
dva@ece.gatech.edu

SEE THE FUTURE

CREATE YOUR OWN