

User-Specific Training of a Music Search Engine

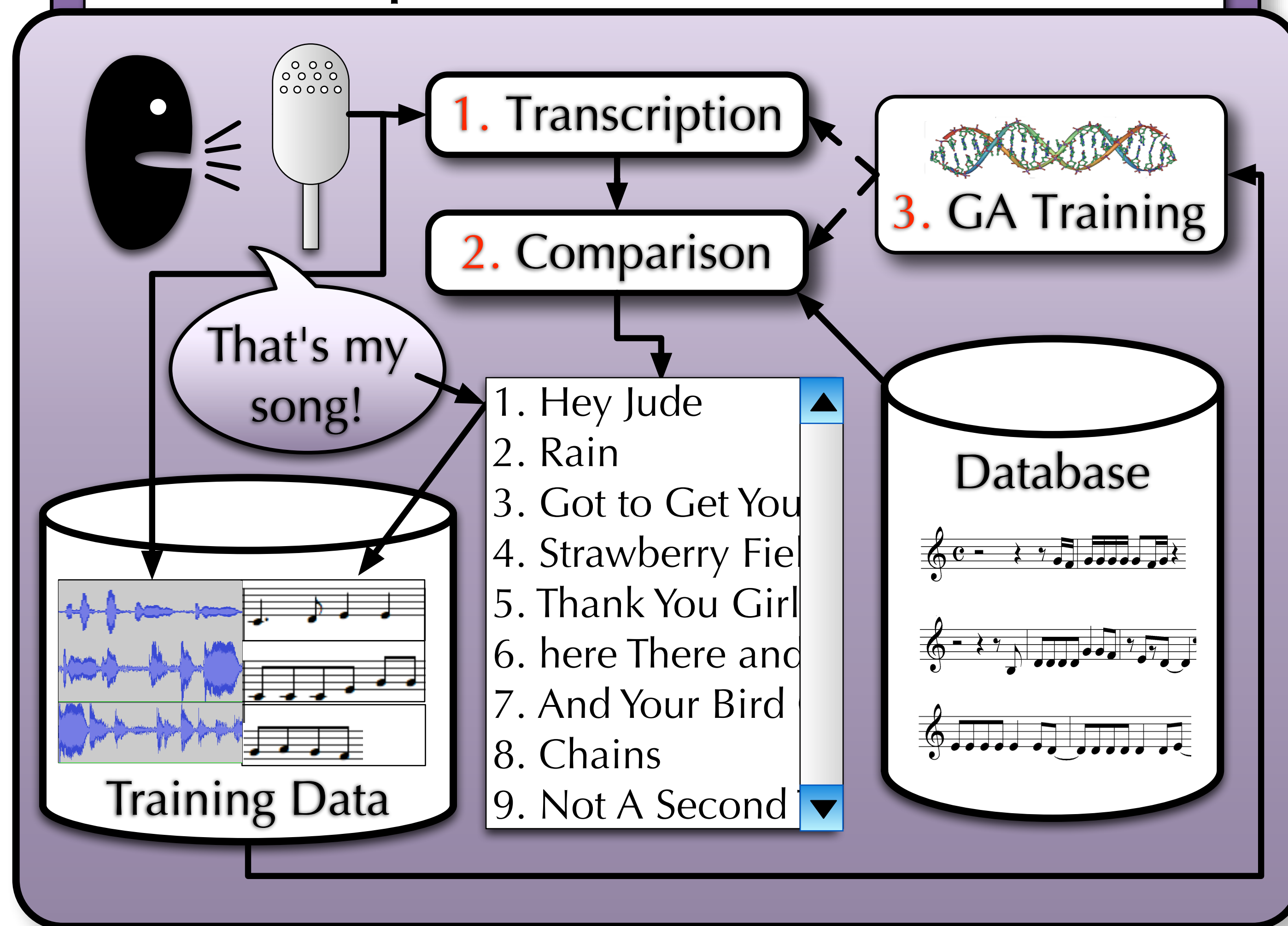
David Little David Raffensperger Bryan Pardo
{d-little,d-raffensperger,pardo}@northwestern.edu



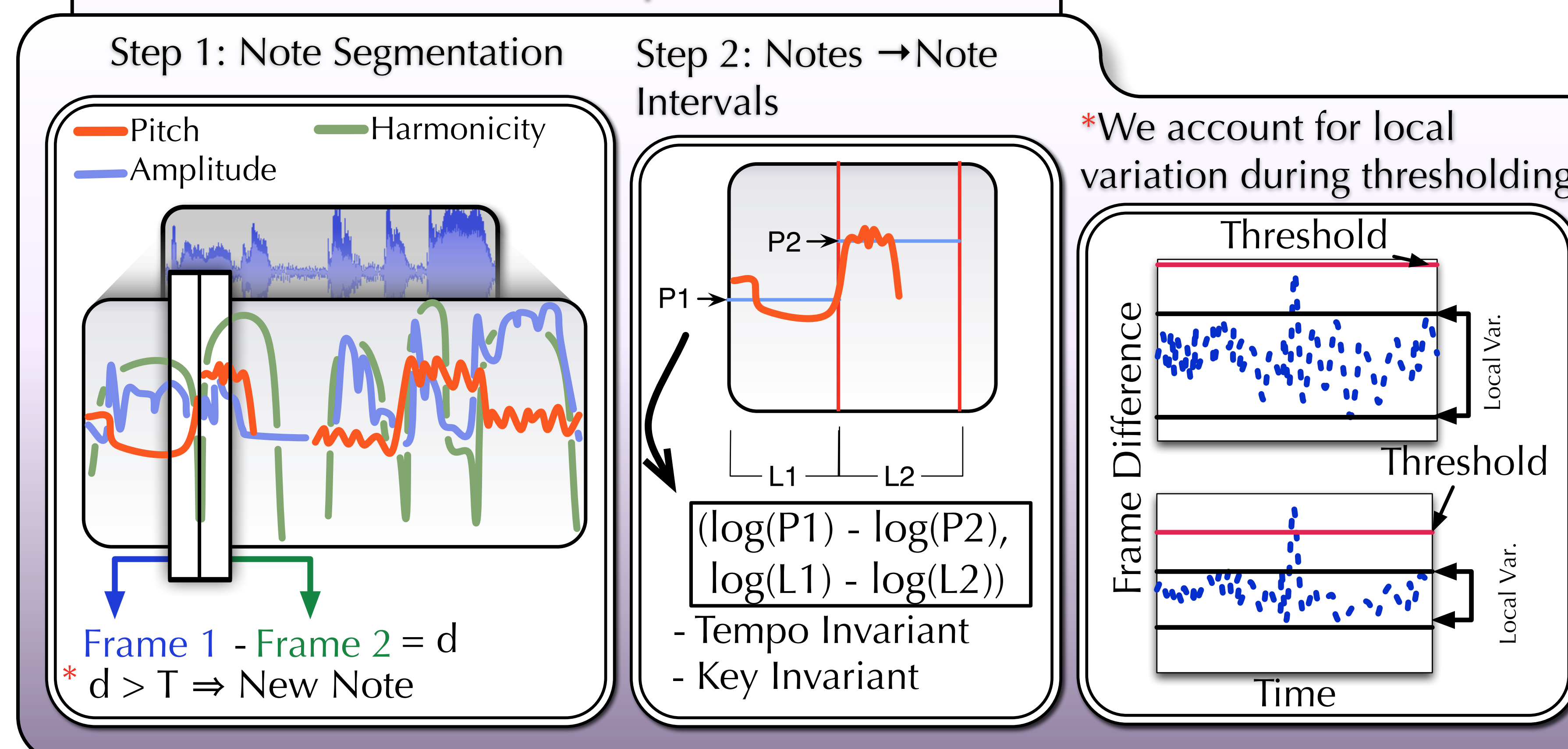
NORTHWESTERN
UNIVERSITY

System Overview

Our system searches a database of songs based on a melody sung into a microphone. Although previous systems have used training before deployment, our system is capable of improving search results after deployment. Users indicate which song they were searching for in the results of their search; this response is used to improve results on future queries. Since training is performed after deployment the system can be tailored to specific users.



1. Transcription

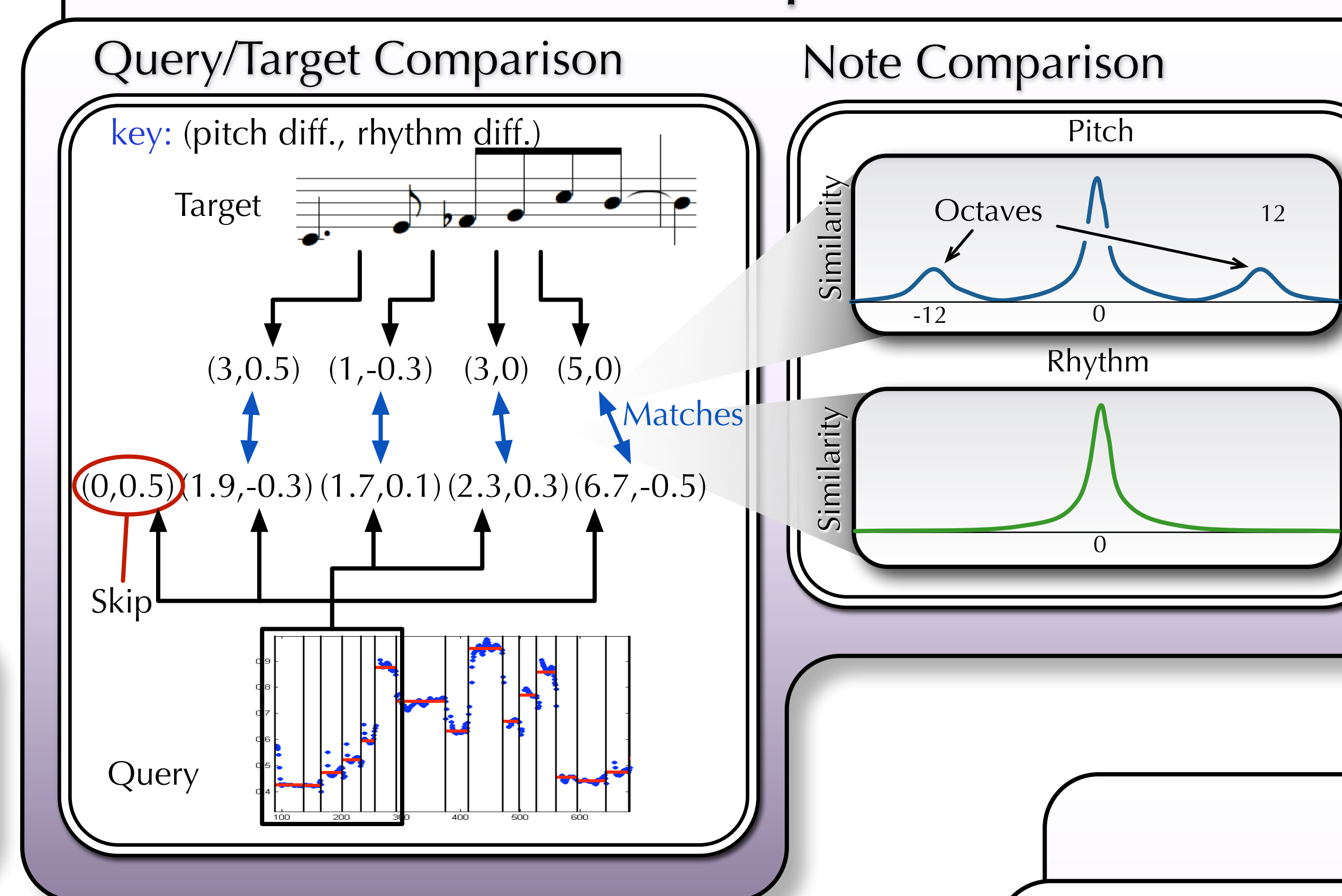


A user's query is first transcribed into a note interval representation. To accomplish this, the pitch, harmonicity and amplitude of each 10ms slice are estimated, and significant changes in these values are used to identify new notes. Our interval representation uses a difference of logs, rather than just a difference, as this has been shown to perform better.

Parameters tuned by GA:

- Pitch Weight
- Harmonicity Weight
- Amplitude Weight
- Threshold

2. Comparison

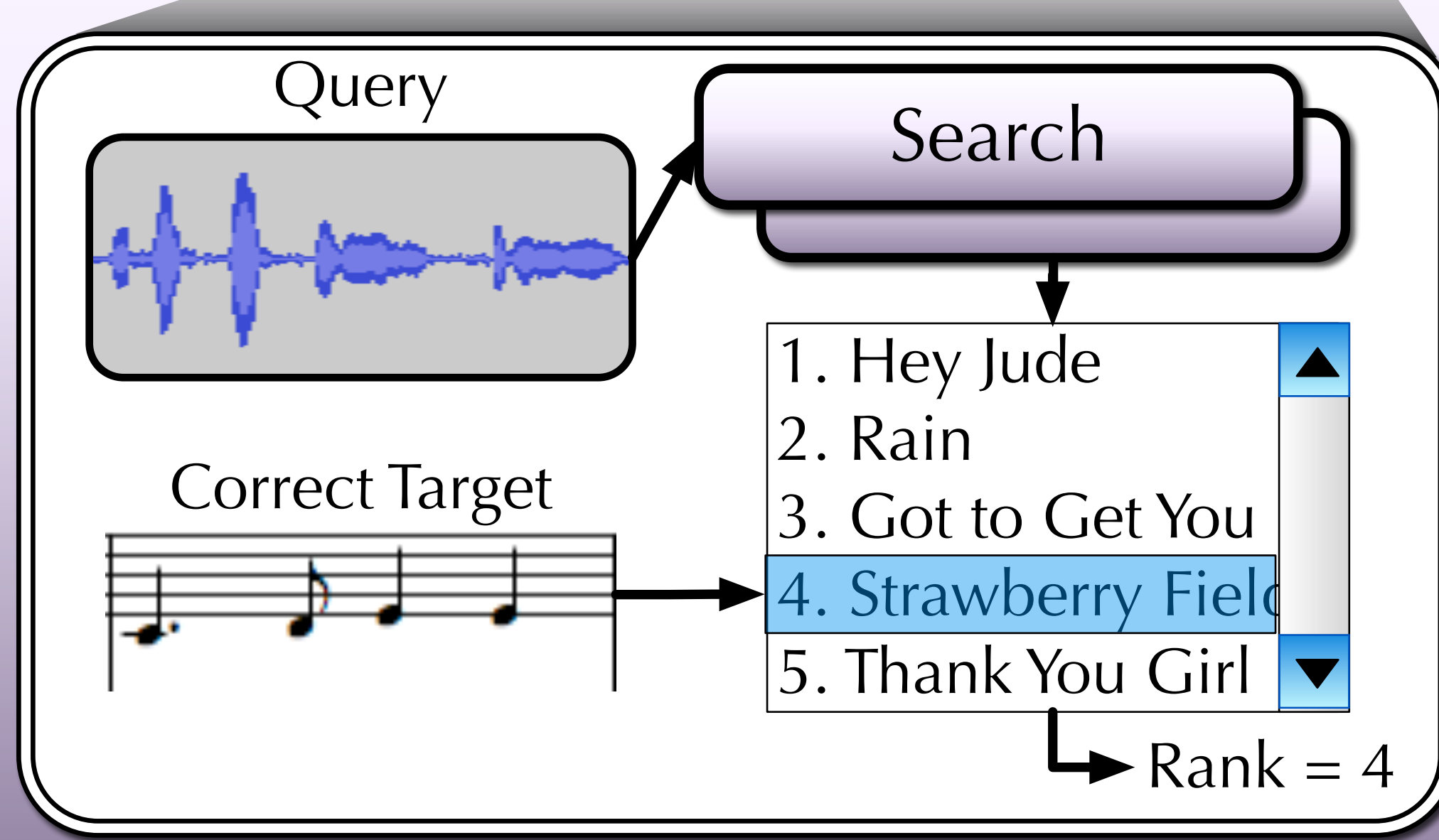
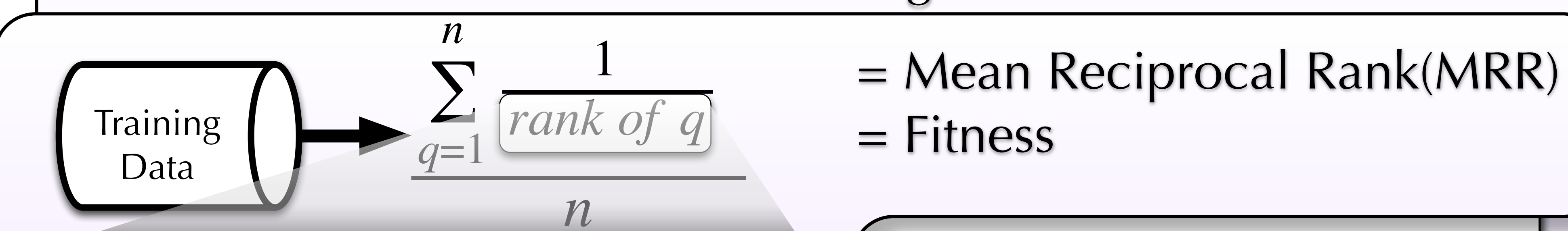


Once transcribed, a query is compared to each song (target) in the database. This is done using a probabilistic string matching technique. We define a simple gaussian based function to determine how similar two note intervals are to each other. This allows inexact matches between intervals (e.g. 2.8 half steps is similar to 2.6 steps, so the two might be matched).

Parameters tuned by GA:

- Rhythm Standard Deviation
- Pitch Standard Deviation
- Octave Decay

3. GA Training

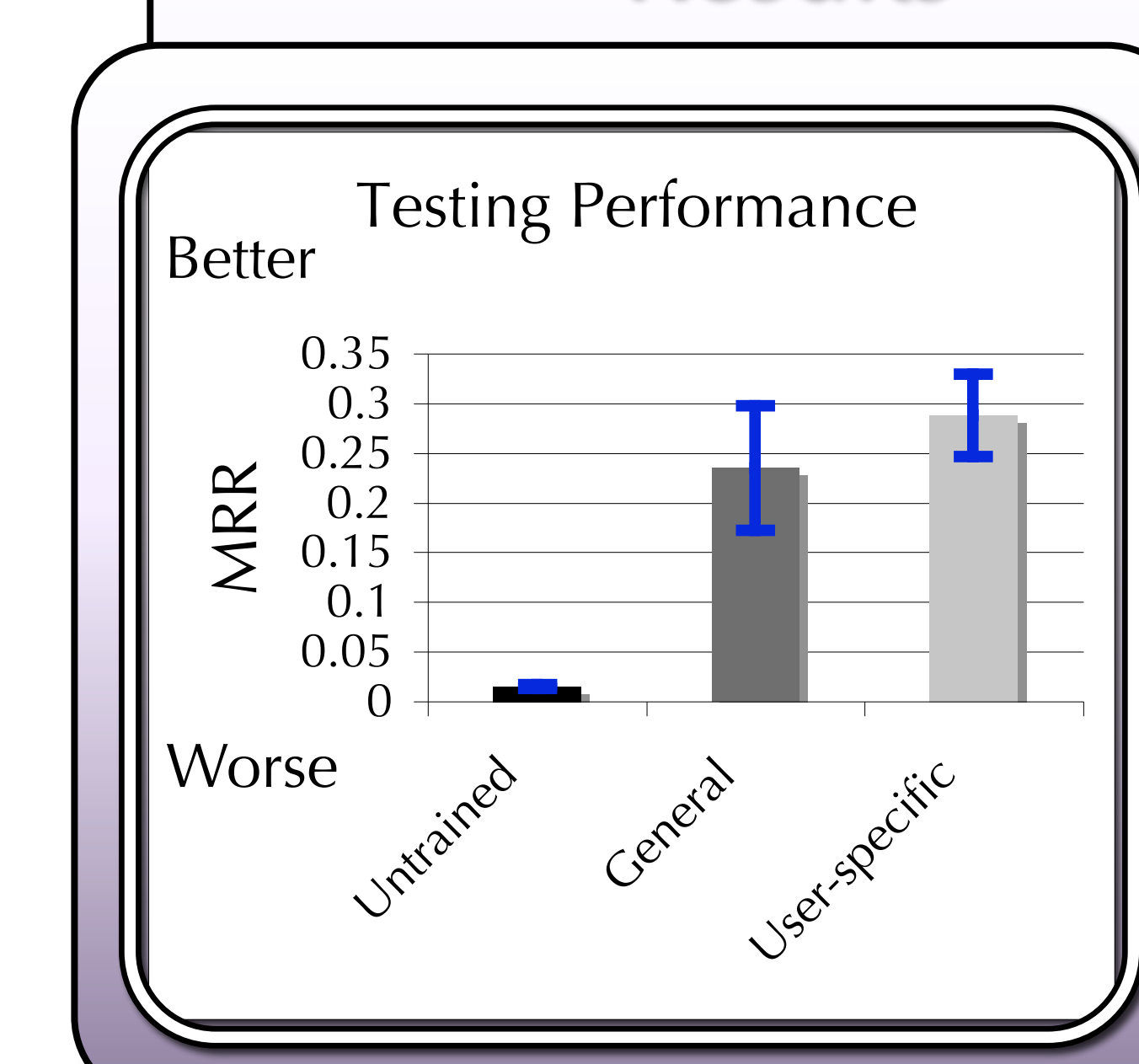


The **seven tunable parameters** our search engine uses for transcription and comparison of melodies are adjusted using a genetic algorithm. For each set of parameters, every query from the training data is searched for using these parameters and the rank of the correct target for that query is found. The mean reciprocal rank (MRR) is then used to determine the parameter set's fitness.

GA Details

- Parameter Representation:** 8 bit binary fraction for each parameter.
- Generations:** 40 (pilot tests showed no further improvement, after 30 generations).
- Population Size:** 60
- Mutation Probability:** 0.02
- Cross-over** is allowed between parameters.

Results



Test data included 100 sung queries (10 queries by 10 singers). Searches were performed on a database of 483 songs. K-fold validation was used to train and test parameters. All conditions were significantly different ($p \leq 0.024$).