

---

# Music 209

## Advanced Topics in Computer Music

### Lecture 3 – Speech Synthesis

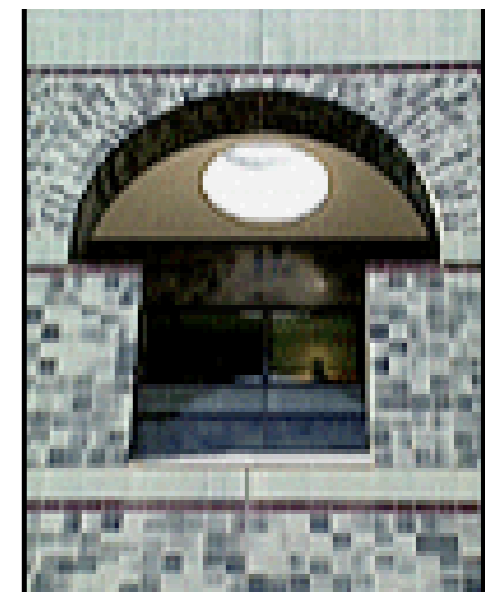
---



**Special guest:  
Robert Eklund**



**2006-2-2**



**Professor David Wessel (with John Lazzaro)**  
([cnmat.berkeley.edu/~wessel](http://cnmat.berkeley.edu/~wessel), [www.cs.berkeley.edu/~lazzaro](http://www.cs.berkeley.edu/~lazzaro))

---

[www.cs.berkeley.edu/~lazzaro/class/music209](http://www.cs.berkeley.edu/~lazzaro/class/music209)

---



# Musical topics for today ...

---

- \* **Pop music lead vocals: a composite of many performances.**
- \* **Note-level concatenative singing synthesis**
- \* **Phrase concatenative synthesis, choirs**
- \* **Project ideas**

# Pop Vocals: Recorded in Isolation Booths

Large-diaphragm  
condenser microphone

Pop shield

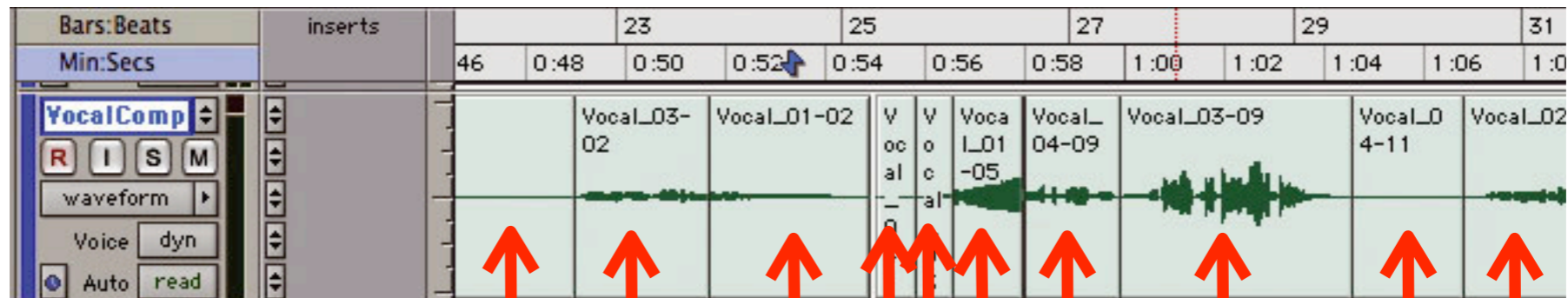
Monitor backing tracks  
via sealed headphones



**Goal: Print a dry vocal with no "room" sound**  
**Dynamic-range management is usually only effect printed.**

# Pop Vocals: Assembled from 'Takes'

Final  
Vocal



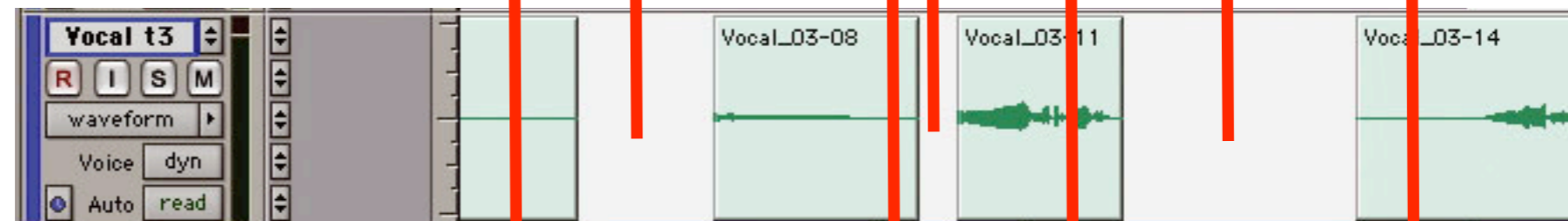
Take 1



Take 2



Take 3



Take 4



# Best take isn't in tune? Pitch correction.



Before

After

Before

After

Cher effect: Play

# Set levels so voice “sits” well in mix

Yellow line is engineer manually moving fader ...



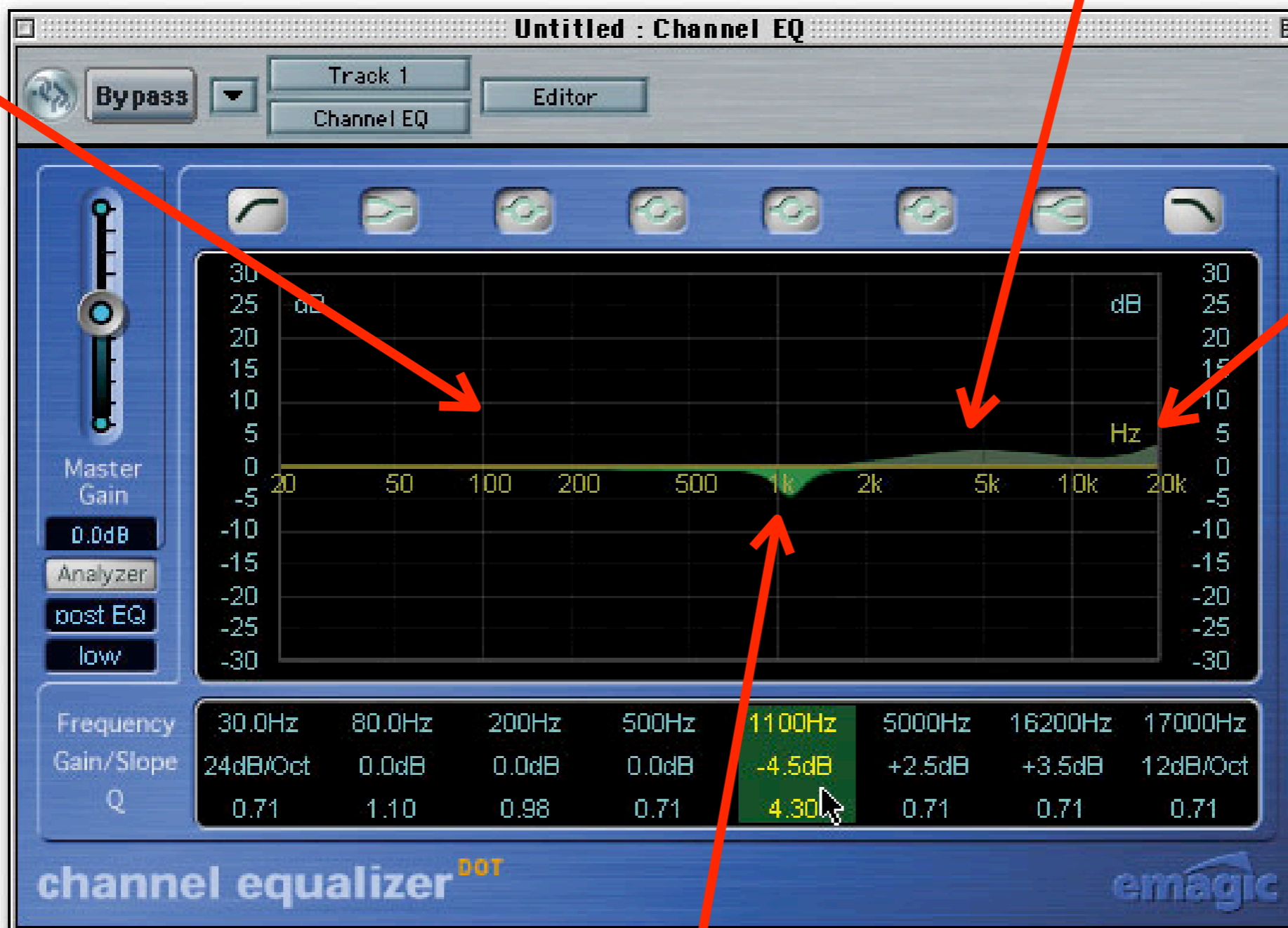
Waveform shows effect of moderate compression.

# EQ to fine tune vocal timbre ...

200 Hz boost/cut - add "warmth" or fix "chestiness"

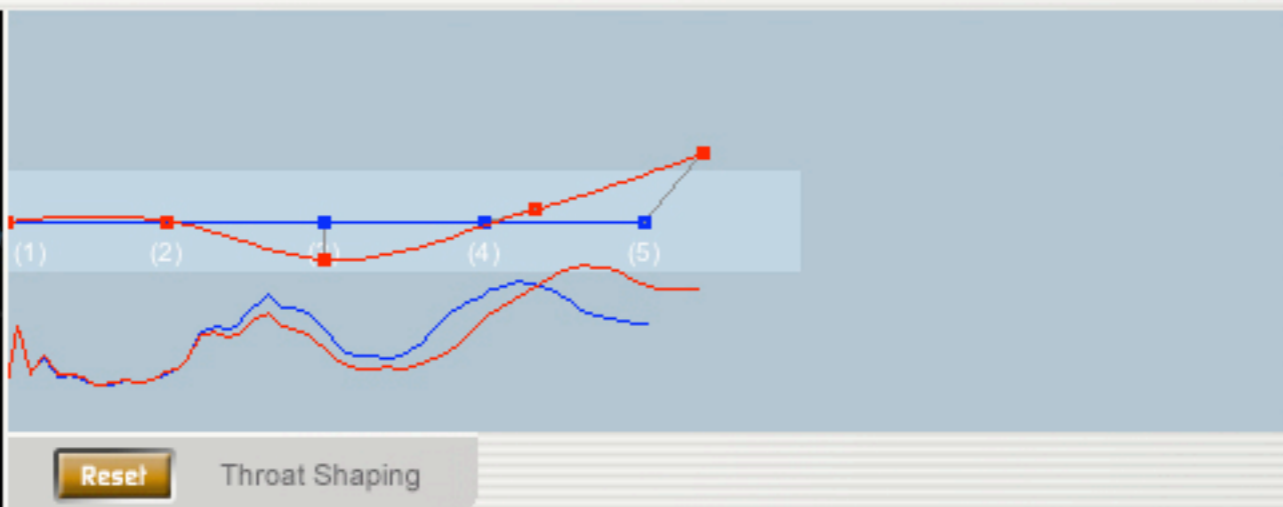
4-6 kHz boost - "Presence"

15 kHz boost - "Air"



Narrow notch cuts to fix timbre "defects" (nasality, etc)

# Voice modeling: 'Physical' modification



Before

After

## THROAT Physical Modeling Vocal Designer

Vocal Range: **alto/tenor** Range

Source Glottal Waveform: **medium** Voice Type

Source Throat: **extreme** Precision

Add Breathiness: **0** Mix, **4000** HP freq (Hz)

Model Throat: **1.00** Length, **1.00** Width

Model Glottal: **100** Pulse Width, **soft** Voice Type

Output: **0** Gain

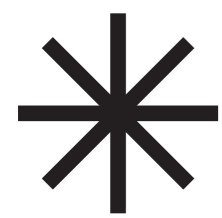
Level Matching: **Off**

Bypass: **Off**

antares

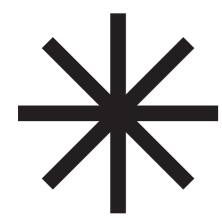


# Reverb: Placing the vocal in a space.



**NOT trying to place all instruments on the record in the same space.**

**Some instruments are totally dry (example: bass drum).**



**Goal is to build a space that works well for the singer and the song.**

**Newest technique: vocal reverb whose character changes line by line, to accentuate words.**



Is this level of  
perfectionism really  
needed for record to be  
commercially successful?

Jagged Little Pill, Alanis Morissette. Released 1995.

Copies Sold: 30 million+. On the short list of best selling albums of all time.

Songs written in the studio in 13 days. As songs were written, they were recorded, and the lead vocals and backing tracks appear on the record as they were originally recorded.



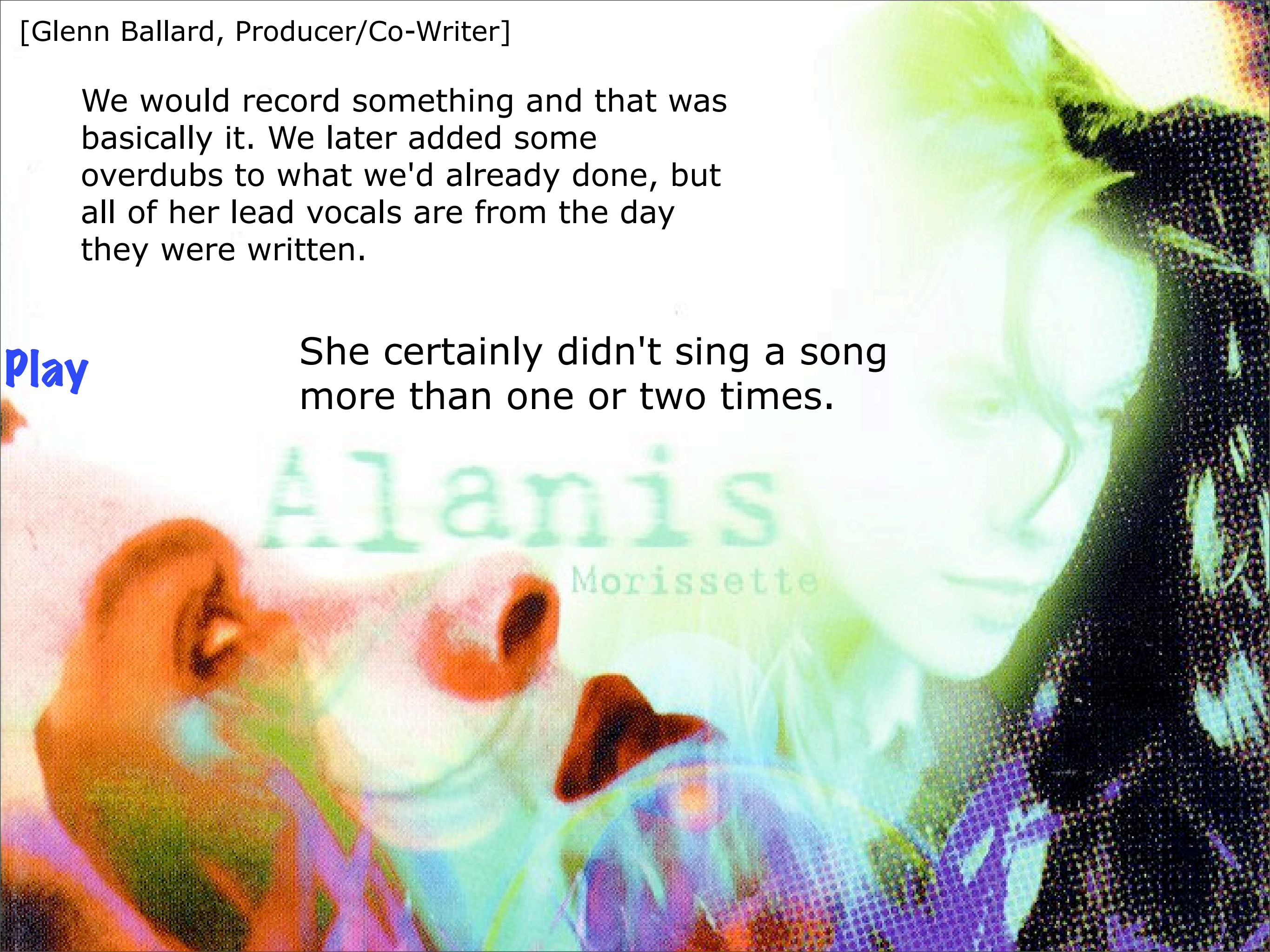
[Glenn Ballard, Producer/Co-Writer]

We would record something and that was basically it. We later added some overdubs to what we'd already done, but all of her lead vocals are from the day they were written.

Play

She certainly didn't sing a song more than one or two times.

Alanis  
Morissette





Music  
Technology  
Group



CREATING 'KANDO' TOGETHER

# Singing Synthesis

---



**Barcelona-Yamaha collaboration began in 2000.  
First VoiceFonts released by Zero-G in Fall 2003.  
Still in early-adopter phase.**



# Vocaloid: Building the database

---

- \* **Concatenative vocal synthesis.** Each virtual vocalist is a sampled human vocalist.
- \* Human vocalist sings from scores with **lyrics of nonsense words** that cover the space of **phonemic and pitch transitions**.
- \* Segmented into **diphones**, converted to a **Fourier** representation, **cleaned of vibrato** and **pitch-bend** in an Auto-Tune-like process.
- \* Phrasing, pitch-bend, vibrato mannerisms of singer captured separately as control data.

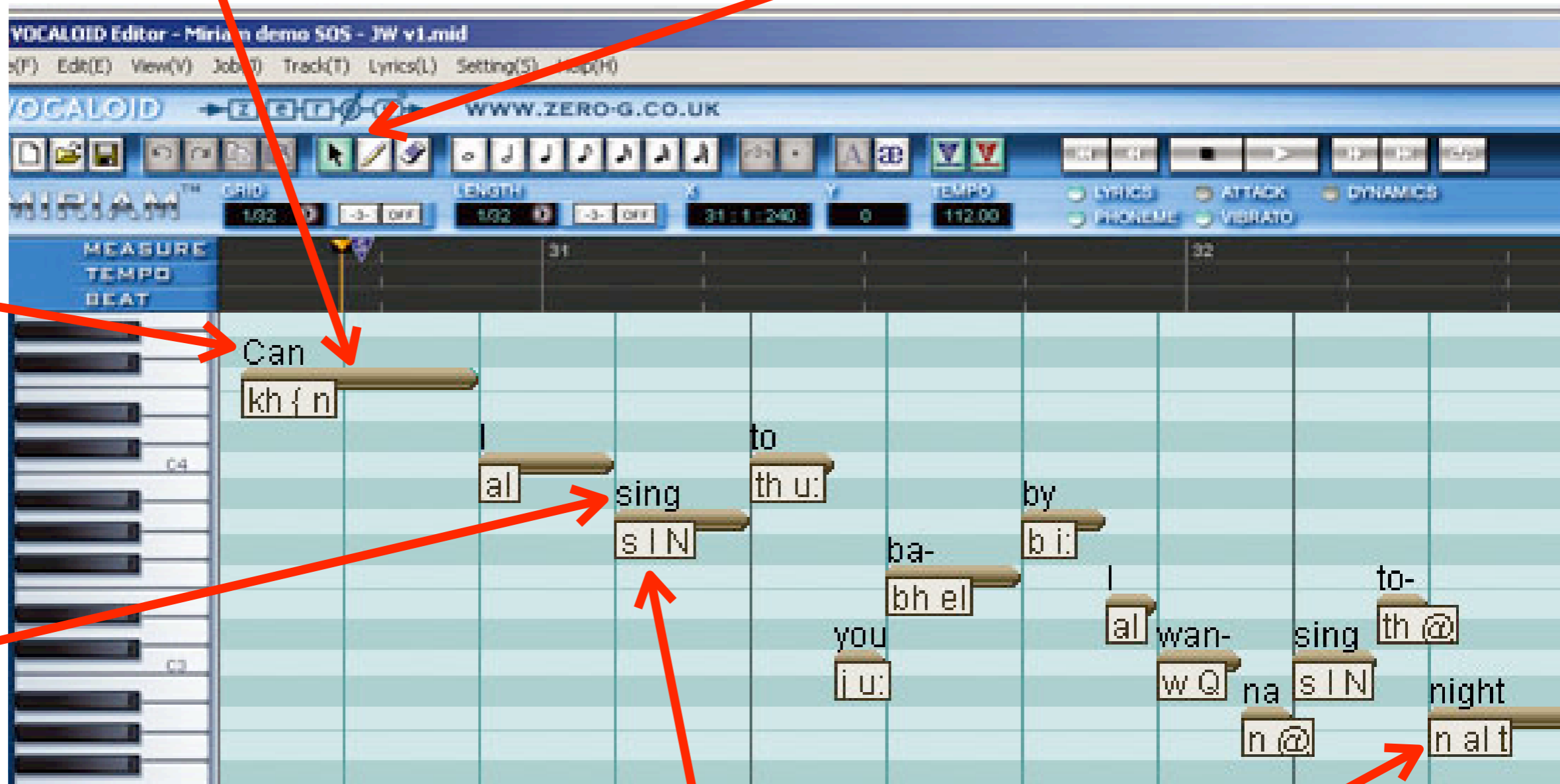
**One virtual vocalist: 500MB to 2.5 GB of data.**



# Vocaloid: Synthesis User Interface

User draws in melody line with a pencil (or import a MIDI file).

User labels each note with a lyric word.



System generates phonemes labels automatically





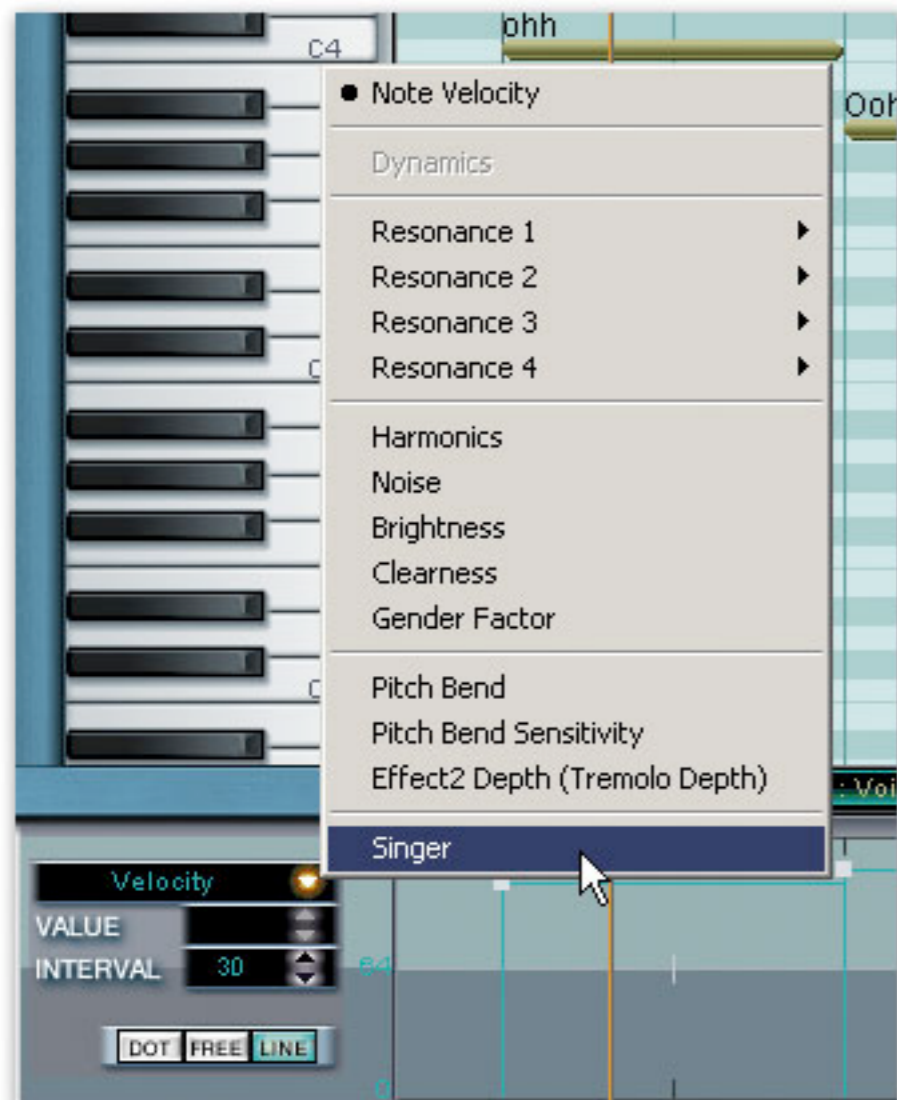
Notate score with icons to humanize performance:

Articulation, legato, vibrato, dynamics, ...

The musical score notation includes the following lyrics and performance elements:

- Lyrics:** Can you sing to me by your side, I want to sing to you tonight.
- Performance Elements:**
  - ATTACK:** Tenuto lines above the notes for 'Can', 'sing', 'to', 'ba-', 'sing', 'to-', and 'night'.
  - VIBRATO:** Wavy lines below the notes for 'sing', 'ba-', 'sing', and 'night'.
  - DYN.:** Dynamic markings 'mf' and 'mp' in boxes below the notes.
  - CRESC.:** Wedge-shaped icons indicating dynamics for 'sing', 'ba-', 'sing', and 'night'.
  - DIM.:** Wedge-shaped icons indicating dynamics for 'to-' and 'night'.





**Many  
continuous  
parameters  
may be drawn  
in by hand ....**



**Can also hand-edit: phonemes, dictionary,  
and raw resynthesis parameters.**

# How does it sound?

---

**The hardest test: Classic songs in English made famous by great singers.**

**Somewhere Over The Rainbow:**

**Play**

**Scarborough Fair**

**Play**



# Easier: Songs written for Vocaloid

---

**I Want a Dog: Written for a  
Canadian TV children's show.**

**Play**

**Your Fish Tank: Novelty Song.**

**Play**



# Yet Easier: Language Unknown to Audience

---

**Japanese song #1**

**Play**

**Japanese song #2**

**Play**



# Other easy cases ...

---

**Background Vocals (Lead  
Vocal is a human singer)**

**Play**

**Scat Singing**

**Play**



# Biggest downsides ...

---

 **Editing takes too long if the goal is realistic results: similar to violin concatenative synthesis.**

 **Using it with a real-time controller has big obstacles: algorithms require lookahead to work well.**

# Voice Project Idea #1

---



# Glossolalia Singing Synthesis ...

 [Sign in / create account](#)

[article](#)

[discussion](#)

[edit this page](#)

[history](#)

*Your continued donations help Wikipedia grow!*

## Glossolalia

From Wikipedia, the free encyclopedia

**Glossolalia** (from the Greek, "γλώσσα" (glossa), tongue and "λαλώ" (lalô), to speak) comprises the utterance of what appears (to the casual listener) either as an unknown foreign language (**xenoglossia**), simply nonsense syllables, or utterance of an unknown **mystical language**; the utterances sometimes occur as part of religious worship (**religious glossolalia**).

**Skeptics** dismiss these cases as simply being in a state of trance, self-hypnotism or religious ecstasy. It is notable that in **Charismatic/Pentecostal Churches** there is often a state of heightened **emotionalism** which may, in the view of skeptics, itself give rise to instances of glossolalia (what Christians in those churches often refer to as **feeling the spirit**).

### Contents [\[hide\]](#)

- [1 Linguistic view](#)
- [2 Psychological view](#)
- [3 Christian view of speaking in tongues](#)
  - [3.1 Context and terms](#)
  - [3.2 Tongues in the New Testament](#)
  - [3.3 Contemporary Christian glossolalia](#)
- [4 Non-Christian glossolalia](#)
- [5 See also](#)
- [6 External links](#)
- [7 Biblical references to speaking in tongues](#)

**Play**



**WIKIPEDIA**  
*The Free Encyclopedia*

### navigation

- [Main Page](#)
- [Community Portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random article](#)
- [Help](#)
- [Contact us](#)
- [Donations](#)

### search

### toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)



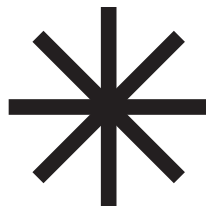
# A good match to concatenation ...

---

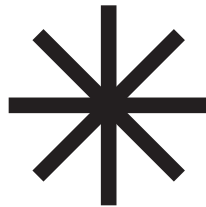
- \* We can design the language with phonemic transitions that sound good.**
- \* There are no native listeners, so no one will hear marginal transitions as synthetic.**
- \* If we let lyrics be generated algorithmically, playing the voice from a MIDI controller becomes possible.**

# Two ways artists approach glossolalia

---

 **Scientifically.** (example: Elizabeth Frazier, of the Cocteau Twins). A linguist, she designs syntax and semantics for a novel language, then writes lyrics in it.

**Project idea: computer tools to help the design process, perhaps with the goal of making concatenative singing synthesis sound good. in the language (Adrian Freed's idea).**

 **Improvisationally.** (example: Lisa Gerrard, of Dead Can Dance).

**Project idea: Sample her a cappella Glossolalia singing, and use it in a concatenative system.**



**Recall: Construct database of complete musical phrases that are browsed via GUI (example: Liquid Saxophone).**

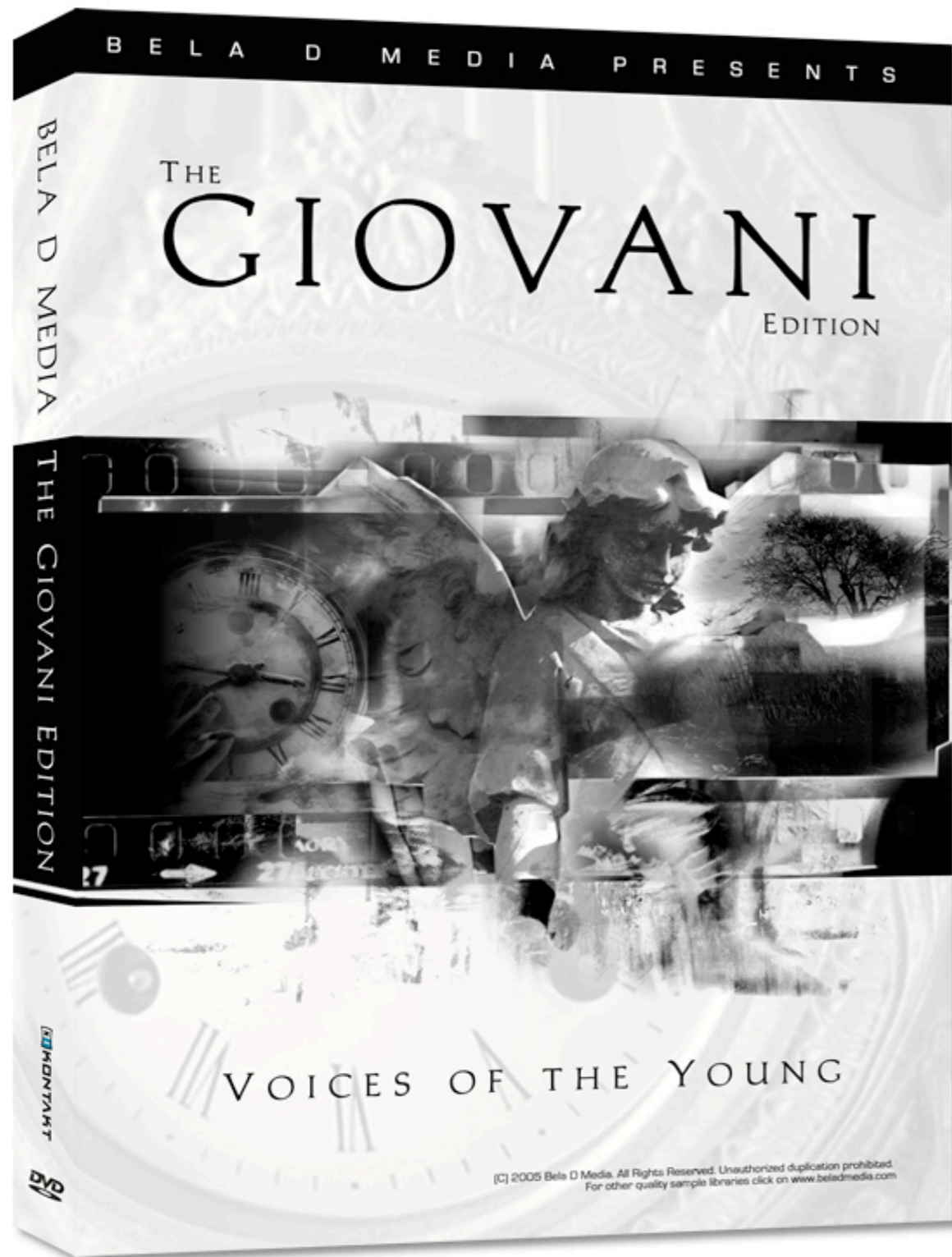
# Phrase-Based Synthesis

---

**Main Problem: Choosing lyrics that would be useful ....**



# Children's choir: \$375. Sold out first run quickly.



## Sampled Latin

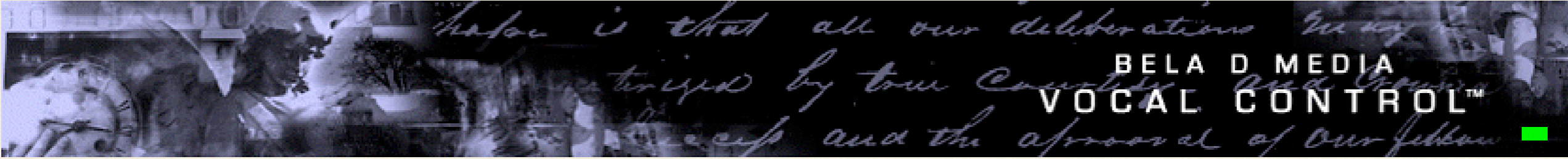
- Agnus Dei
- Benedictus
- Dies Irae
- Veritas Domini
- Morte Aeterna
- Peccata Mundi
- Requiem Aeternam

**Play**

# Rudimentary phrase concatenation ...

**Vocal Control - Untitled** [Window Title Bar]

File Midi VC Pool Help [Menu Bar]



**PHRASE 1 [E2]**

ahg	nus	day	yee	ben	neh	dek	tus
dee	yas	ee	ray				

Link OFF

**PHRASE 2 [F2]**

ver	ree	tas	doh	mee	nee		

Link OFF

**PHRASE 3 [G2]**

ben	neh	neh	tus	day			

Link OFF

**PHRASE 4 [A2]**

dek	tus	doh	mee	nee			

**GLOBAL**

BYPASS OFF	<input checked="" type="radio"/> CYCLE PHRASE	LOCK [C2]	<input type="radio"/> FINGERED	LEGATO [D2]
CLEAR	<input type="radio"/> PHRASE 1 TIME		<input checked="" type="radio"/> MOMENTARY	
UNDO CLEAR	<input type="radio"/> LOCK ELEMENT		<input type="radio"/> TIMED ---->	500 ms

**Boys Ensemble**

ahg	mee
nus	nee
day	
yee	
ben	
neh	
dek	
tus	
dee	
yas	
ee	
ray	
ver	
ree	
tas	
doh	

OVERWRITE

# Harder to do with pop music choirs ...

---

The Voice Vol. 1 features 300 verbal vocal phrases between 2 and 8 bars focused mainly on pop, dance and RnB productions. All vocal phrases can be combined with each other.

The verbal phrases include: "listen 2 the groove", "keep me movin on", "liftin me higher", "party everybody", "ready 4 my luv", "u make me wanna dance", "universal love", "feel so high", "sexy dancer", "when will u stop playing" and many more.



# Voice Project Idea #2

---



# There has to be a better way ...

---

The verbal phrases include: "listen 2 the groove", "keep me movin on", "liftin me higher", "party everybody", "ready 4 my luv", "u make me wanna dance", "universal love", "feel so high", "sexy dancer", "when will u stop playing" and many more.

**Project idea: Come up with a principled idea for creating a useful phrase library (words and melody + signal processing) that is data driven from lyric and MIDI databases on the web.**

**Project Proposals Due March 1, 11:59 PM,  
via email to David and John ... see website.**

