



Visual-Voice TalkSync™ is a lip-synching plugin for Kinetix/Discreet 3D Studio MAX™. It analyzes .wav files and automatically makes a user-defined mouth speak the words by inserting the appropriate keyframes in the 3D Studio MAX track editor. Articulation includes individual vowels and consonants.

The heart of the system is a proprietary voice recognizer that analyzes the acoustics of the sound to determine the positions that the lips, jaw (teeth) and tongue would assume to produce the sounds. (Most commercial recognizers do statistical analysis of audio files to recognize words and even phrases.) This makes the resulting animated motion extremely accurate, even for gibberish, and eliminates the need to pre-train the recognizer for each speaker's voice.

The mouth topology that can be controlled is extremely flexible, so mouths of all types can be animated, including bird bills, robots and any other user-defined shapes an animator might want to create. Initial setup of the mouth model is quick and easy- Just determination of key animation parameters. There is no need to create individual mouth positions for each phoneme. A selection of sample mouths, complete with preset motion parameters, suitable for commercial use is included on the CD.

## Mouth Creation and Setting Animation Parameters

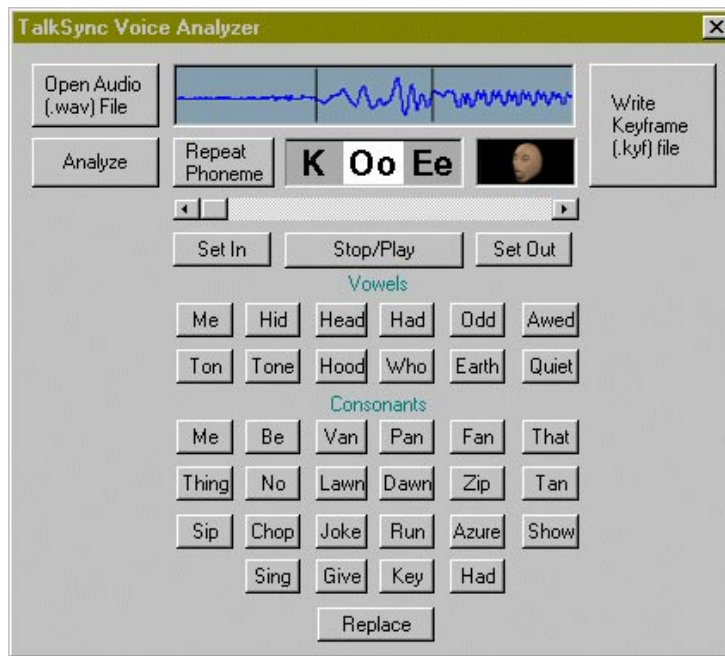
Mouths can have three elements: lips, a jaw (teeth) and a tongue. Each is animated separately: The lips and tongue animate by morphing from position to position. The jaw animates by swinging downward around a pivot point and x-y translation. Not all elements are required: Teeth that ride up and down with the lips can be created as part of the lips and the jaw can be left out, etc. There are no restrictions on the shapes of any of these elements: They can be very realistic or take on completely artificial shapes.



TalkSync will automatically interpolate lip, tongue and jaw position from in between a range of user-defined 'extreme' speech positions: These are the extreme vowels: 'Ah' (Mouth wide open), 'Ee' (Lips spread) and 'Oo' (Lips rounded into a small circle), as well as distinctive consonants: 'F/V' (Lower lip pulled in to touch upper teeth), 'B/P' (Puckered lip closure), 'Sh/Ch/Zh' (Lips puckered open), as well as distinctive tongue positions: Touching the roof of the mouth behind the teeth as if to make the 'L' sound, between the teeth as if to make the 'Th' sound, and withdrawn into the back of the mouth to make the 'G' sound.

# Speech Recognition and Mouth Animation

TalkSync consists of two modules: the Voice Analyzer and the Keyframe Writer.



*Voice Analyzer User-Interface*

The **Voice Analyzer** is a standalone application that can be operated on any NT or Windows 95/98 workstation. (It is *not* necessary to have 3D Studio MAX installed on this machine.) It accepts audio .wav files and generates a .kyf file which holds the data of how the mouth should move. It does not need to be pre-trained to the speaker because it operates on acoustics rather than statistical word-matching.

The **Keyframe Writer** is a plugin for 3D Studio MAX that takes the mouth movement data in the .kyf file and applies it to the appropriate mouth in a scene. It automatically inserts the appropriate keyframes into the 3D Studio MAX track view to make the selected elements mouth the words. These keyframes can be edited with all the normal 3D Studio MAX editing tools, so it's easy to customize the speech.

## Product Status

TalkSync is expected to be available in the first quarter of 2004.

For more information, please contact:

Visual-Voice, Inc.  
1486 Beacon St.  
Brookline, MA 02446  
617-277-4317  
Fax: 617-232-8748  
www.visual-voice.com  
sales@visual-voice.com