We previously reported a face detection system based on color segmentation using HSV. It was shown that the color is more effective than other colors not only in accurate segmentation but also in effective extraction of facial features. The first is crucial for detection and the latter for recognition. When it comes to video footages of news program, sound often accompanies the video and persons express themselves by moving facial parts while speaking. In this paper we improve the face detection in speed using both sound and video in a combined way. First, the rate of syllables spoken is estimated from the sound. Next, for a beginning short video clip of each new scene, a differential image is formed with the frame distance corresponding to the rate to find mouth and eyes. This enables us to reduce the number of sampling points for segmentation to a great degree and to enhance the reliability of the detection. Also music is discriminated from speaking by the estimation. These contribute to much faster detection of face.