

# SONG PROMPTER: AN ACCOMPANIMENT SYSTEM BASED ON THE AUTOMATIC ALIGNMENT OF LYRICS AND CHORDS TO AUDIO

Matthias Mauch    Hiromasa Fujihara    Masataka Goto

National Institute of Advanced Industrial Science and Technology (AIST), Japan

{m.mauch, h.fujihara, m.goto}@aist.go.jp

## 1 Introduction

We present *Song Prompter*, a software system that acts as a performance guide by showing horizontally scrolling lyrics, chords, beats marks and bar marks in a graphical user interface, together with an audio accompaniment consisting of bass and MIDI drums. A song outline displays the song structure, including names and positions of sections for easy overview and navigation. *Song Prompter* enables users to sing and play live along the timeline of an original song, without having to memorize lyrics and chords or turning pages. Chord labels, and bass and audio playback can be transposed to a different key, and the playback speed can be changed.

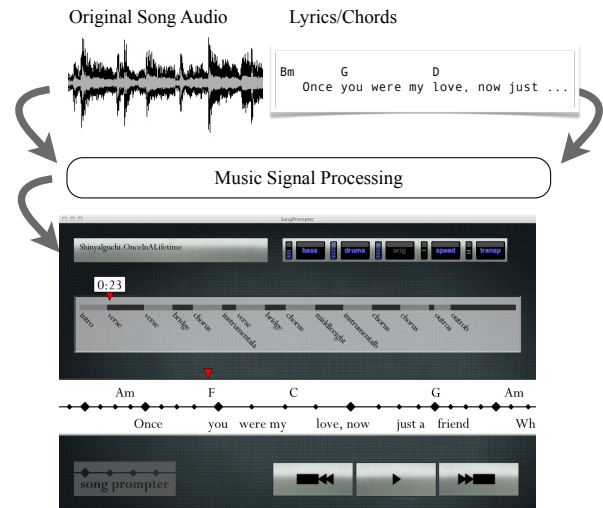
Unlike existing score following systems (for a review see [1]), karaoke systems, or musical computer games such as *Rock Star* and *Guitar Hero*, *Song Prompter* requires no musical score input, and chords and lyrics need not be manually aligned with the audio. Instead, we use an automatic lyrics-to-audio alignment method [4], which requires only the original audio recording of a song and a text file with lyrics and chords in a format similar to that often found on the internet. The drum and bass accompaniment is also extracted automatically from the original audio recording. Therefore, the music selection is not limited by commercially available transcriptions.

## 2 Usage Scenarios

Many usage scenarios are conceivable. For example, in a cover band rehearsal situation, a band member can propose a new song by his favourite artist, and the band can immediately start playing the song based on the lyrics, beats and chord progression. When a band performs live, *Song Prompter* can literally act as an automatic prompter. In a more informal setting, it can be used when a party of friends want to sing together, and someone has brought along a guitar: *Song Prompter* is more convenient than song books because no pages need to be turned and everyone always knows the correct song position. Since the text is scrolling past, it can be displayed in a larger font than is possible in a static book format.

## 3 Music Signal Processing

All processing is done prior to playback. The display requires the physical onset times of chords, words, sections, beats and bars in a song. The Internet-style chords and lyrics representation of the song is automatically parsed, assuming that section names are also provided in this representation. At this point, chords, words and section names are not associated with their physical onset time. We define a left-to-right hidden Markov model (HMM) of chord



**Figure.** System outline. For a demo see <http://www.youtube.com/user/SongPrompter>

and phoneme states, with joint chroma and MFCC observations. Chroma features are extracted from the corresponding original audio file, and MFCC features from the re-synthesized melody line segregated from the audio [3]. Offline Viterbi alignment then estimates the physical times of every chord, word and section. The beat marks and larger bar marks complete the white song stream pane (see Figure). Their positions are obtained using [2]. The fundamental frequency and the amplitude of the partials in the bass line are estimated using PreFEst [3].

At performance time the bass line is re-synthesized on the fly, using the frame-wise parameter estimates. Bass drum, snare drum and hi-hat MIDI notes are triggered, based on the extracted beat and bar times. The playback speed can be changed, and bass and original audio can be transposed.

This research was supported by CrestMuse, CREST, JST.

## 4 References

- [1] Roger B. Dannenberg and Christopher Raphael. Music score alignment and computer accompaniment. *Communications of the ACM*, 49(8):38–43, 2006.
- [2] Matthew E. P. Davies, Mark D. Plumbley, and Douglas Eck. Towards a musical beat emphasis function. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009)*, 2009.
- [3] Masataka Goto. A real-time music scene description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals. *Speech Communication (ISCA Journal)*, 43(4):311–329, 2004.
- [4] Matthias Mauch, Hiromasa Fujihara, and Masataka Goto. Lyrics-to-audio alignment and phrase-level segmentation using incomplete internet-style chord annotations. In *Proceedings of the 7th Sound and Music Computing Conference (SMC 2010)*, 2010.