

【発明の名称】

画像認識を利用して音声認識を行う方法及び装置(METHOD AND APPARATUS FOR VOICE RECOGNITION USING VIDEO RECOGNITION)

【要約】

本発明は、画像認識に基づいて、音声認識の開始と終了をより正確に認識することができる音声認識方法及び装置に関し、本発明の実施例に係る音声認識方法は、音声認識モードの切り換え前に、第 1 ビデオデータ、または、第 1 オーディオデータのうち少なくとも一つに基づいて発話の開始を判断するステップと、発話の開始と判断した場合には、音声認識モードに切り換えて、ユーザーの音声コマンドを含む第 2 のオーディオデータを生成するステップと、音声認識モードに切り換えた後の第 2 のビデオデータまたは前記第 2 のオーディオデータのうち少なくとも一つに基づいて、発話の終了を判断するステップを含むことを特徴とする。本発明によると、ユーザーは別のジェスチャーを示す必要なく、唇を動かすだけで音声認識機能を実行することができ、画像認識を介して音声認識の開始と終了をより明確に判断することができる効果がある。

【請求項 1】

電子機器が音声命令を認識する方法において、音声認識モードの切り換え前の第 1 ビデオデータ、または、第 1 オーディオデータのうち少なくとも一つに基づいて発話の開始を判断するステップと、発話の開始と判断した場合には、音声認識モードに切り換えて、ユーザーの音声コマンドを含む第 2 のオーディオデータを生成するステップと、音声認識モードに切り換えた後の第 2 のビデオデータまたは前記第 2 のオーディオデータのうち少なくとも一つに基づいて、発話の終了を判断するステップと、を含むことを特徴とする認識方法。