

VOICE CONVERSION: STATE OF THE ART AND PERSPECTIVES

Yannis Stylianou

University of Crete, Grece

RESUMEN

Voice Transformation refers to the various modifications one may apply to the sound produced by a person, speaking or singing. In other words, Voice Transformation aims at the control of non-linguistic information of speech signals such as voice quality and voice individuality. Voice Transformation covers a wide area of research from speech production modeling and understanding to perception of speech, from natural language processing, modeling and control of speaking style, to pattern recognition and statistical signal processing.

Voice Transformation was considered as a hot, novel and fast-growing topic in 1990s having as potential application the concatenated speech synthesis systems where new (virtual or target) voices could be created without requiring to pass through the quite expensive process of developing new voices. By that time, it was widely accepted that Voice Transformation systems were far from providing the required performance. With the recent developments in speech synthesis this need is more pronounced. There is an increasing demand for high quality Voice Transformation methods not only for creating target or virtual voices, but also to model various effects (e.g., Lombard effect), synthesize emotions, to make more natural the dialog systems which use speech synthesis etc.

In this talk I will review the state-of-the-art Voice Transformation methodology showing its limitations in producing good speech quality and its current challenges. Addressing quality issues of current voice transformation algorithms in conjunction with properties of the speech production and speech perception systems I will try to pave the way for more natural Voice Transformation algorithms in the future. Facing the challenges, it will allow Voice Transformation systems to be applied in important and versatile areas of speech technology. Besides speech synthesis, Voice Transformation has other potential applications in areas like entertainment, film, and music industry, toys, chat rooms and games, dialog systems, security and speaker individuality for interpreting telephony, high-end hearing aids, vocal pathology and voice restoration.

Yannis Stylianou Recibió el Diploma en Ingeniería Eléctrica en 1991 y el MSc y PhD en Procesado de Señal en la ENST de Paris, Francia, en 1992 y 1996 respectivamente. Desde 1996 hasta 2001 trabajó en AT&T Lab. Research (NJ, USA). En 2001 ingresó en los laboratorios Bell (Lucent Technologies) en NJ (USA) (ahora Alcatel-Lucent). Desde 2002 trabaja como Profesor Asociado en la Universidad de Creta, en el Departamento de Ciencias de la Computación, y como Investigador Asociado en el Laboratorio de Redes de Telecomunicaciones del Instituto de Ciencias de la Computación.

Es miembro del Comité Técnico del IEEE Speech and Language. Es editor asociado del EURASIP Journal on Speech, Audio and Music Processing y de las EURASIP Research Letters in Signal Processing, y vice-chair de la Acción COST 2103: "Advanced Voice Function Assessment". Fué editor asociado para la revista IEEE Signal Processing Letters y estuvo en comité de gestión (MC) de la Action COST 277: "Nonlinear Speech Processing". Entre otros proyectos en el FP6, participó en la Red de Excelencia SIMILAR coordinando la tarea de fusión de las modalidades de voz y escritura. Tiene 9 patentes y es miembro de IEEE y de la Technical Chamber of Greece.