



جامعة الموصل

كلية الهندسة

استخدام المنطق المضرب والشبكات العصبية لمعالجة الإشارة الصوتية وتطبيقاتها

ندى نويل ابراهيم

رسالة ماجستير علوم في هندسة الحاسوب

بإشراف

الأستاذ الدكتور باسل شكر محمود

كانون الثاني/ ٢٠١٧م

ربيع الثاني/ ١٤٣٨هـ

المستخلص بلغة الرسالة

استرجاع الإشارة الكلامية المشوهة (Speech Restoration) هو أحد أهم تطبيقات معالجة إشارة الكلام؛ لأنه يمكن السامع من الاستماع إلى الإشارة المشوهة وفهمها بصورة واضحة، حيث يتم تشويهِ إشارة الكلام لأسباب كثيرة وبطرق عديدة.

في هذه الدراسة تم معالجة حالة تشويه الإشارة الكلامية؛ نتيجة تغيير مدى الترددات الناتج عن تغيير فترة الصوت (Time Shrinking)، مثل: إعادة تشغيل الصوت بمسجل ذي سرعة أكبر؛ أي: زيادة سرعة الكلام.

تم تصميم وتنفيذ نظامين لاسترجاع الإشارة الكلامية المشوهة، ويتكون كل نظام من ثلاث مراحل رئيسية تشمل: مرحلة استخراج المميزات (Feature Extraction)، مرحلة عمليات المعالجة (Processing Operation) التي يتم إجراؤها على المميزات المستخرجة، وأخيراً مرحلة استخلاص الموجة الصوتية من المميزات المعدلة (Speech Mel Frequency Cepstral Coefficient (MFCC)). يتشابه النظامان في المرحلتين الأولى والأخيرة، حيث تم استخدام تقنية معاملات $\text{Mel Frequency Cepstral Coefficient}$ لاستخراج المميزات المهمة من الإشارة الكلامية في المرحلة الأولى، ومعكوس معاملات $\text{Inverse Mel Frequency Cepstral Coefficient}$ (IMFCC) لاستخلاص الموجة الصوتية من معاملات $\text{Mel Frequency Cepstral Coefficient}$ (MFCC) للتردد الطيفي في المرحلة الأخيرة من أجل الاستماع إلى الإشارة الكلامية والتأكد من صحتها. أما مرحلة المعالجة (المرحلة الثانية) للنظام الأول فقد تمت باستخدام الشبكات العصبية (Neural Network (NN)، وكذلك تمت للنظام الثاني عن طريق إضافة المنطق المضيف إلى النظام ودمج تقنيتي المنطق المضيف والشبكات العصبية وتكوين شبكة عصبية مضببة (Fuzzy Neural Network (FNN).

ولقياس كفاءة النظامين تم حساب نسبة أعلى إشارة إلى الضوضاء ((PSNR Peak Signal to Noise Ratio) وقيمة متوسط مربع الخطأ (Mean Square Error (MSE)) لقيم الإشارة الأصلية التي تم تسجيلها وإشارة المسترجعة التي تم الحصول عليها بعد عمليات المعالجة التي أجريت في النظامين، وكان متوسط قيم (PSNR) لخمس كلمات يساوي (26.1256) للنظام الأول، و (26.0863) للنظام الثاني، ومتوسط قيمة (MSE) أيضاً لخمس كلمات يساوي (0.0546) للنظام الأول، و (0.0505) للنظام الثاني.

ويدعم استرجاع الإشارة الكلامية المشوهة نتيجة تغيير مدى تردداتها عدة تطبيقات مهمة، مثل: استرجاع الإشارة المشوهة من التسجيلات المحطمة أو التي تعمل بسرعة تختلف عن سرعة الإشارة الأصلية، وكذلك استرجاع الإشارة المشوهة نتيجة مرورها في أوساط تختلف في كثافتها.

د. عمر موفق محمود اليوسف
معاون العميد للشؤون العلمية
والدراسات العليا

Abstract

Speech restoration is one of the most important processes application in speech processing domain because it enables a listener to listen and understand the distorted speech signal clearly. The speech signal may be distorted for many reasons and in many ways. This study handle a distorted speech signal that was shrunk to its half period time (just like when the speed of a play back recorder increases to its double normal speed).

Two systems were designed to restore the distorted speech signal, each system consists of three stages (Feature extraction, Processing operation, and Speech waveform reconstruction). The first and last stage are the same, whereas Mel Frequency Cepstral Coefficient (MFCC) was used for Feature Extraction, and Inverse Mel Frequency Cepstral Coefficient (IMFCC) was used for speech waveform reconstruction. In the second stage, neural network and fuzzy neural network were used for the first and second system respectively.

The peak signal to noise ratio (PSNR) and the mean square error (MSE) were calculated according to the recorded and the restored speech, the obtained results of 50 words for PSNR was (26.1206), (26.0863) and for MSE was (0.0546), (0.0505) for the first and second system respectively.

University of Mosul
College of Engineering



Neuro-Fuzzy based processing of speech and its applications

Nada Noel Ibrahim

M.Sc. Thesis

Computer Engineering

Supervised by

Professor

Dr. Basil Sh. Mahmood

January/2017 A.D.

Raby al-Thany/1438 A.H.