

Digital Audio Signal Processing for Time Stretching and Pitch Shifting

Diese Arbeit untersucht und vergleicht Algorithmen für die Tonhöhen- und die Wiedergabegeschwindigkeitsänderungen von Audiosignalen. Beim Ändern der Dauer durch Streckung oder Stauchung der Signalmesswerte im Zeitbereich wird die Tonhöhe (Pitch) beeinflusst, und umgekehrt. Die untersuchten Algorithmen sollen die Modifikationen der Tonhöhe und der Dauer voneinander entkoppeln.

Es wurden die Algorithmen TD-PSOLA, WORLD, LPCNet und CLPCNet untersucht. TD-PSOLA teilt das Audiosignal in zeitliche Intervalle auf, modifiziert diese und setzt sie wieder zusammen. Die anderen Algorithmen analysieren das Signal, um die Grundfrequenz und Spektralhülle zu extrahieren, wodurch das Sprachsignal rekonstruiert wird. LPCNet und CLPCNet basieren auf neuronalen Netzwerken, welche bei der Synthese auf Basis vorhergehender Ausgaben neue Vorhersagen bestimmen.

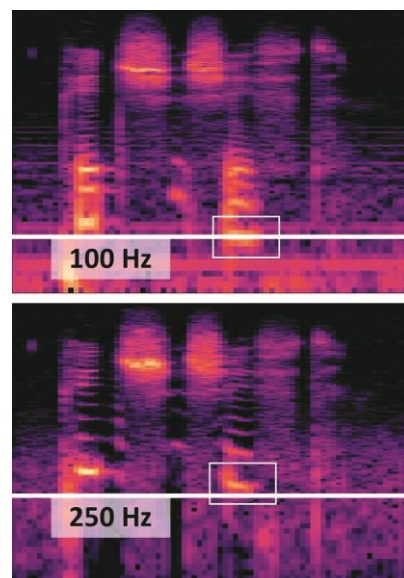
Anhand subjektiver Kriterien sowie technischer Limitationen der Algorithmen wurden Grenzwerte für die maximalen Pitch- und Zeit-Modifikationen mit subjektiv akzeptabler Qualität definiert, sowie deren Robustheit gegenüber Rauschen getestet. Die Algorithmen wurden zudem durch Fehlermasse, die die Pitch-Abweichung bestimmen, objektiv verglichen. In einer durchgeführten Umfrage wurden ergänzend die Ergebnisse der Algorithmen, bei Zeit- und Pitch-Modifikationen, einander gegenübergestellt. Um die Unterschiede der neuronalen Netzwerke ausschliesslich bezüglich ihrer Architektur bewerten zu können, wurden beide Netzwerke mit demselben Datensatz neu trainiert.

Die Ergebnisse aus der Umfrage waren in den Auswertungen mit den objektiven Fehlermassen wiederzufinden. Bei einer Erhöhung der Wiedergabegeschwindigkeit waren die Unterschiede zwischen den Algorithmen klein. Bei Streckungen hingegen lieferte TD-PSOLA die subjektiv beste Qualität, bis zu einer Verdoppelung der Laufzeit, was ebenfalls durch die Umfrage bestätigt wurde. Bei grösseren Streckungen lieferte LPCNet die subjektiv besten Resultate, auch bei Audiosignalen mit hohem Rauschanteil. Für Erhöhungen des Pitches lieferten TD-PSOLA und WORLD besonders gute Ergebnisse, sowohl subjektiv als auch in den objektiven Fehlermassen. Bei der Verringerung des Pitches waren die Algorithmen grundsätzlich besser als im Erhöhen, wobei TD-PSOLA und WORLD grundsätzlich die besten Resultate lieferten.



Diplomierende
Simon Brun
Marc Frey

Dozierende
Philipp Matthias Schmid
Sigisbert Wyrsch



Dieses Bild zeigt das Spektrogramm eines Audiosignals (obere Hälfte), dessen Pitch um den Faktor 2.5 erhöht wurde (untere Hälfte). Die weissen Linien dienen zur Hervorhebung des Pitch-Shifts.