

Algorytmy automatyzacji tworzenia baz jednostek akustycznych w syntezie mowy polskiej

dr inż. Janusz Rafałko, Politechnika Warszawska, Wydział MiNI

Temat przedstawia koncepcję systemu komputerowego pozwalającego na zautomatyzowanie procesu oznaczania, segmentacji i wycinania jednostek akustycznych z naturalnego sygnału mowy. W przedstawionym systemie tymi jednostkami są alofony akustyczne. Na ich podstawie tworzone są bazy alofonowe indywidualnych głosów. Bazy takie stosowane są w konkatenacyjnej syntezie mowy TTS. Przygotowanie takich baz w obecnych systemach jest procesem wymagającym profesjonalnej wiedzy i ręcznego podejścia w ich tworzeniu, co jest procesem bardzo czasochłonnym. Przedstawiona metoda w znacznym stopniu automatyzuje ten proces. Wykorzystywany jest do tego zmodyfikowany i dostosowany do potrzeb systemu algorytm DTW. Zestawione są w nim ze sobą dwa sygnały mowy: sygnał syntezowany tworzony na podstawie wzorcowej bazy alofonów i sygnał mowy naturalnej, która poddawana jest segmentacji. Analiza sygnałów mowy odbywa się w dziedzinie częstotliwości. W wyniku działania systemu otrzymywany jest zbiór jednostek, które następnie tworzą bazę alofonową. Odbywa się to w oparciu o opracowane w tym celu algorytmy odrzucania, selekcji i kontroli jakości. Na otrzymanych bazach przeprowadzone zostały odsłuchowe testy jakości, z których wynika, że bazy otrzymane w sposób automatyczny nie odbiegają jakością syntetyzowanej z nich mowy od tej syntetyzowanej przy wykorzystaniu baz wzorcowych.