

Review of Bachelor's Thesis

Student: Gajda Richard
Title: Conversion of Whispered to Normal Voice (id 22505)
Reviewer: Černocký Jan, doc. Dr. Ing., DCGM FIT BUT

- 1. Assignment complexity** **more demanding assignment**
Nadprůměrně náročné zadání výzkumného charakteru, byť bylo postavené na publikovaném článku.
- 2. Completeness of assignment requirements** **assignment fulfilled**
Zadání bylo splněno, kladně hodnotím především práci s daty, kde student provedl velmi kvalitní sběr a přípravu data setu. Provedl rovněž řadu experimentů s konverzí řeči, výsledky jsou uspokojivé pro systém natrénovaný na cílového mluvčího, generalizace je horší, což se dalo předpokládat. Vyhodnocení mohlo být formálnější, např. pomocí speech to distortion ratio (SDR) nebo formálních poslechových testů alespoň s několika posluchači.
- 3. Length of technical report** **in usual extent**
Práce je v obvyklém rozmezí, ale kvalita jednotlivých částí je nevyrovnaná, v popisu extrakce akustických parametrů bych ocenil rigoróznější matematický popis, v experimentální části zase přehled nastavení všech relevantních experimentálních parametrů. Za velmi kvalitní pokládám popis dat a jejich přípravy.
- 4. Presentation level of technical report** **70 p. (C)**
Viz výše, bylo by dobré zpracovat jak na formálnějším popisu jak matematiky, tak experimentů a implementace, práce je v některých částech psána stylem "pohádka", ocenil bych věcnější popis postavený na schématech, tabulkách pseudo-algoritmech, atd.
- 5. Formal aspects of technical report** **85 p. (B)**
Formálně je práce slušně zpracovaná, je logicky členěná a čtivá. Práce je psaná relativně kvalitní angličtinou s minimem chyb. K práci mám řadu drobných poznámek, předám autorovi výtisk, ve kterém řádila oponentská červená fixa.
- 6. Literature usage** **75 p. (C)**
Student důkladně analyzoval hlavní článek pojednávající o konverzi šeptané řeči, u dalších pramenů (vokodéry, neuronové sítě) bych ocenil pečlivější studium. Je škoda, že se od autorů článku nepodařilo získat jejich data-set a srovnat práci s jejich výsledky.
- 7. Implementation results** **85 p. (B)**
Výstupem práce je sada skriptů využívajících knihovny jazyka Python pro strojové učení, a klient-server aplikace pro konverzi hlasu na mobilu. Především ale oceňuji vytvořená data, v práci jsem ale postrádal informaci o tom, zda budou k dispozici.
- 8. Utilizability of results**
Práce má potenciál využití v práci skupiny Speech@FIT, data jsou potenciálně zajímavá pro výzkumnou komunitu. Pokud by se mobilní aplikace dopracovala, má šanci na využití v různých scénářích i na případnou komercializaci.
- 9. Questions for defence**
 - Na schématu 3.2 upřesněte, zda byly pro konverzi skutečně využity statické parametry z neutrální řeči a delta-koeficienty z šeptané?
 - Jak byla vyhodnocována Mel Cepstral Distortion a jak koreluje s poslechovou kvalitou ?
 - Je při pořízení nahrávky na mobilním telefonu tato nějak segmentována (pomocí VAD nebo rovnoměrně) nebo je na server poslána vcelku ?
- 10. Total assessment** **78 p. good (C)**
Technicky a experimentálně slušná práce se zajímavými výsledky a funkční aplikací jedoucí na mobilním telefonu (byť ne nativní, ale klient-server). Kvalitu práce snižuje technická zpráva, která je čtivá, ale postrádá potřebnou technickou přesnost.

In Brno 3 June 2021

Černocký Jan, doc. Dr. Ing.
reviewer