

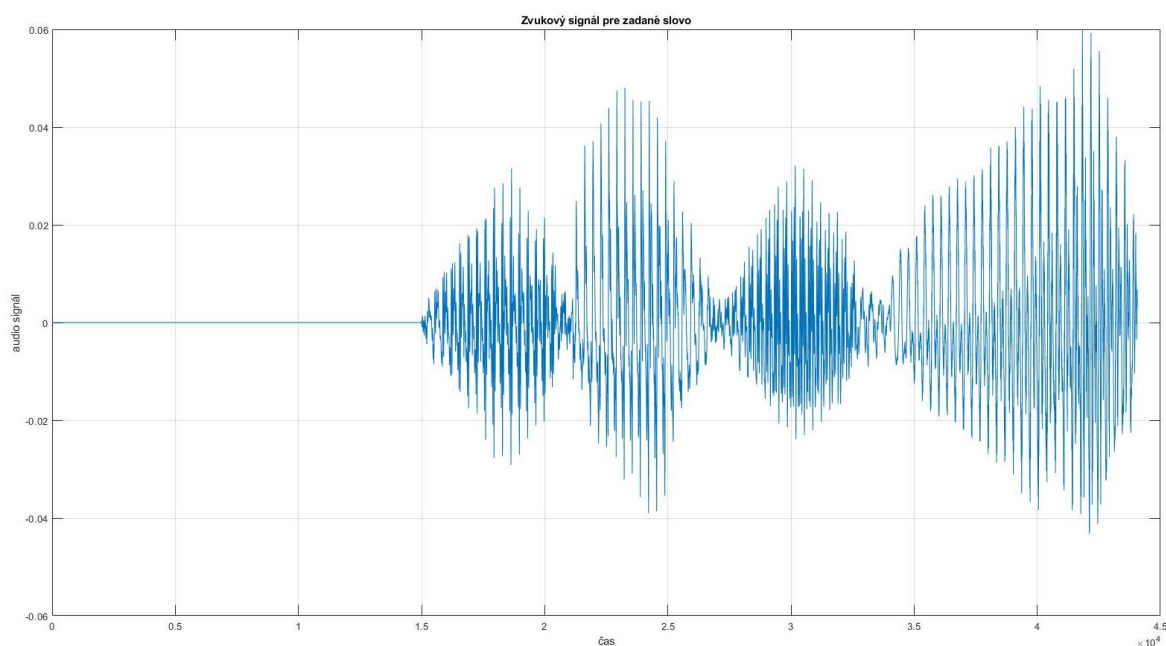
Prevod textu na reč

Tento projekt je rozdelený na 3 návrhy riešenia so stúpajúcou náročnosťou ale so zvyšujúcou kvalitou a prirodzenosťou reči.

V riešení č. 1 som si nahral záznamy fonetickej výslovnosti jednotlivých písmen v abecede. Z týchto nahrávok som musel vyrezať iba žiadaný signál. Keďže tento signál obsahoval šum a závisel od prostredia (veľkosti izby, počtu ľudí v miestnosti) alebo od kvality mikrofónu, nepodarilo sa mi na tento proces vymyslieť nejaký algoritmus a musel som písmená vyseparovať ručne. Ďalším problémom bolo, že nahrávky museli mať približne rovnakú intenzitu hlasu, pretože po poskladaní písmen za seba v tvare nejakého krátkeho slova zneli veľmi neprirodzene.

Riešenie č. 1 (skript solution1) pracuje s užívateľským vstupom. Vstupný reťazec je možný poskladať z klasickej slovenskej abecedy (obsahuje ch, medzeru, neobsahuje písmeno q, dlhé samohlásky ani mäkké na spoluhláskach). Algoritmus prečíta vstup po jednotlivých znakoch a k nim priraduje príslušný zvukový signál. V prvých testovaniach takto zoradené písmená za seba zneli viac ako hláskovanie než prevod na reč.

Na tomto grafe je slovo „ahoj“ zložené zo záznamov jednotlivých písmen A, H, O, J

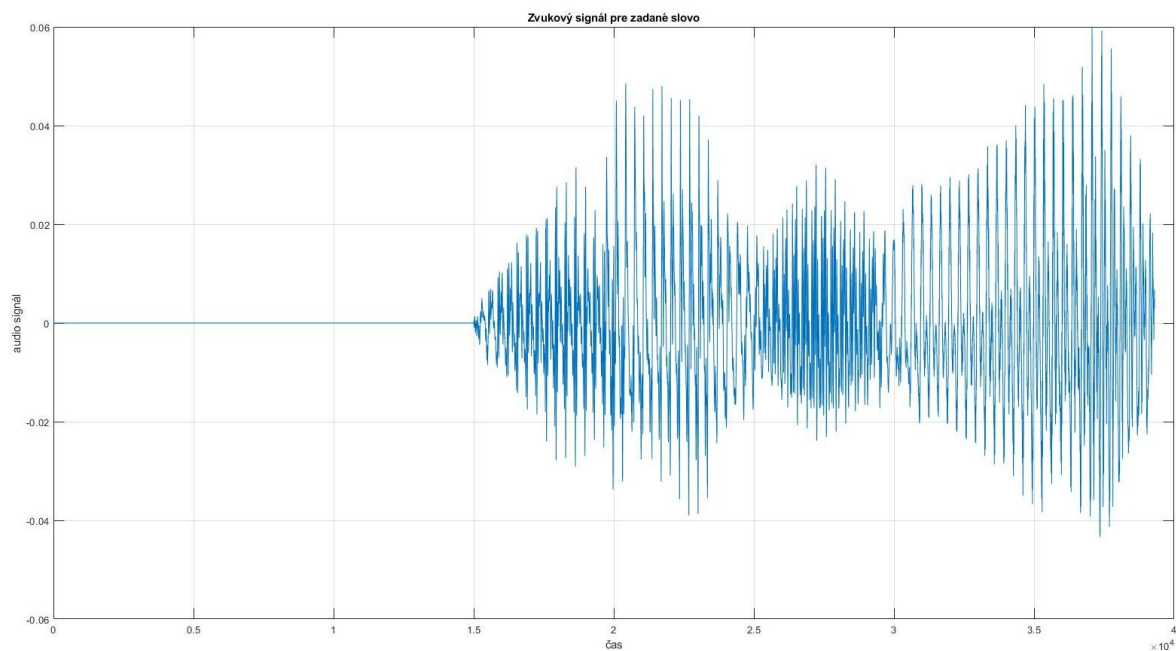


Preto som sa zo signálov nahraných písmen snažil vytvoriť signál čo najviac podobný nahranému slovu. Po testovaní tejto metódy som usúdil, že to nebude možné vytvoriť s týmto datasetom, ktorý mám k dispozícii. Rovnaké písmeno v kombinácii s rozličnými susednými znie inak. Ďalej treba brať do úvahy rozdelenie slov na slabiky a pred touto slabikou umiestniť krátky pokles signálu. Po študovaní gramatického slovníka som zistil, že rozdeľovanie slov na slabiky je dosť komplexné a nebude ho možné naprogramovať v algoritme.

Ďalším nápadom na vylepšenie prirodzenosti reči bola úprava výsledného signálu tak, že medzi signálmi jednotlivých písmen nebude taká veľká medzera, ktorá zapríčiňovala sekanie reči. Do algoritmu v riešení č. 1 som umiestnil parameter overlap, ktorý určuje, koľko percent signálu predchádzajúceho písmena bude prekrytý nadchádzajúcim písmenom.

Nasledujúci graf zobrazuje slovo „ahoj“ so zapnutým parametrom overlap s hodnotou 1/5.

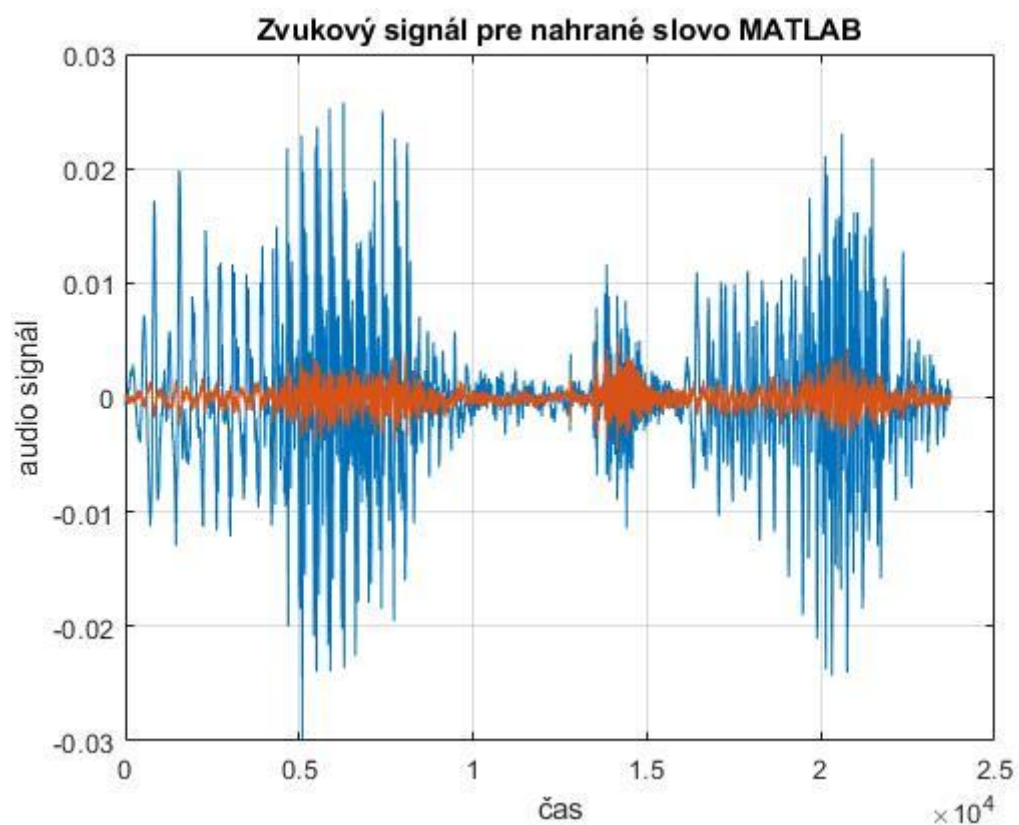
Znamená to, že signál na konci písmena A je sčítaný so signálom začiatku písmena H. Dĺžka tohto intervalu je $\text{overlap} * \text{dĺžka signálu A}$.



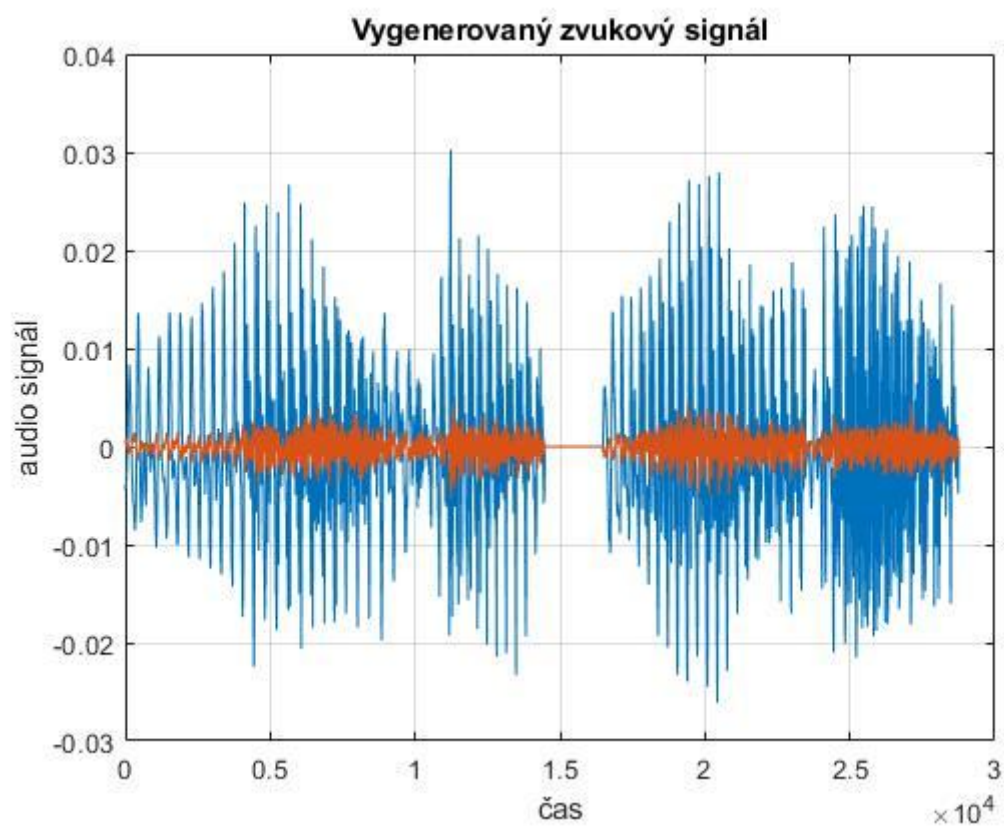
Ale ani s touto úpravou slová neznejú prirodzene. Príklady vytvorených slov sú v priechniku recordings.

V riešení č. 2 som zvolil nový postup. Plán bol nahráť si nejaké náhodné slová a z týchto slov separovať skupiny slov. Čiže slová by som neskladal z jednotlivých písmen ale rozdelil by som zadané slovo na skupiny slov a hľadal zhodu v naučených skupinách slov. Problém bol zase v tom že napr. slabika resp. skupina slov „le“ inak znie na začiatku slova, na konci slova, pred alebo po slabike. V prípade chýbajúcej skupiny slov, ktorú nemá program naučenú, by som dané písmeno nahradil nahraným samostatným písmenom z riešenia č. 1. Pre toto riešenie by som potreboval veľké množstvo nahrávok slov. Preto som zostal iba pri demonštratívnom návrhu tohto riešenia a porovnaním so skutočným signálom. Nahral som slovo „MATLAB“ a toto slovo sa pokúsil zložiť zo slov „mama“, z ktorého som vyseparoval slabiku „ma“ a slova „larva“, z ktorého som použil slabiku „la“. Zvyšné písmená T a B som použil z riešenia č. 1.

Graf pre nahrané slovo „MATLAB“ vyzerá takto:



A graf pre vygenerované slovo touto metódou vyzerá nasledovne:



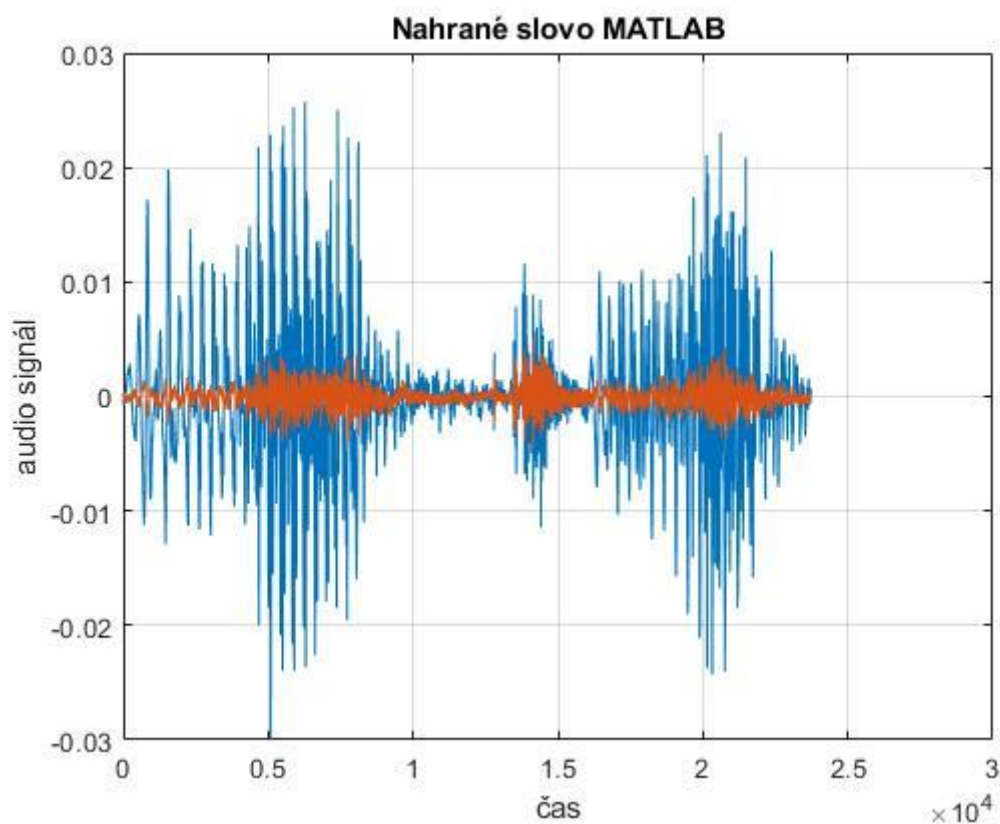
Záznam reči výrazne získal na prirodzenosti a zrozumiteľnosti ale stále sa veľmi nezhoduje s priebehom nahraného signálu.

Preto som pristúpil k návrhu riešenia č. 3, ktoré by počítalo s ešte omnoho väčším počtom záznamov. Zdroje píšú o 10+ hod kvalitných štúdiových nahrávkách.

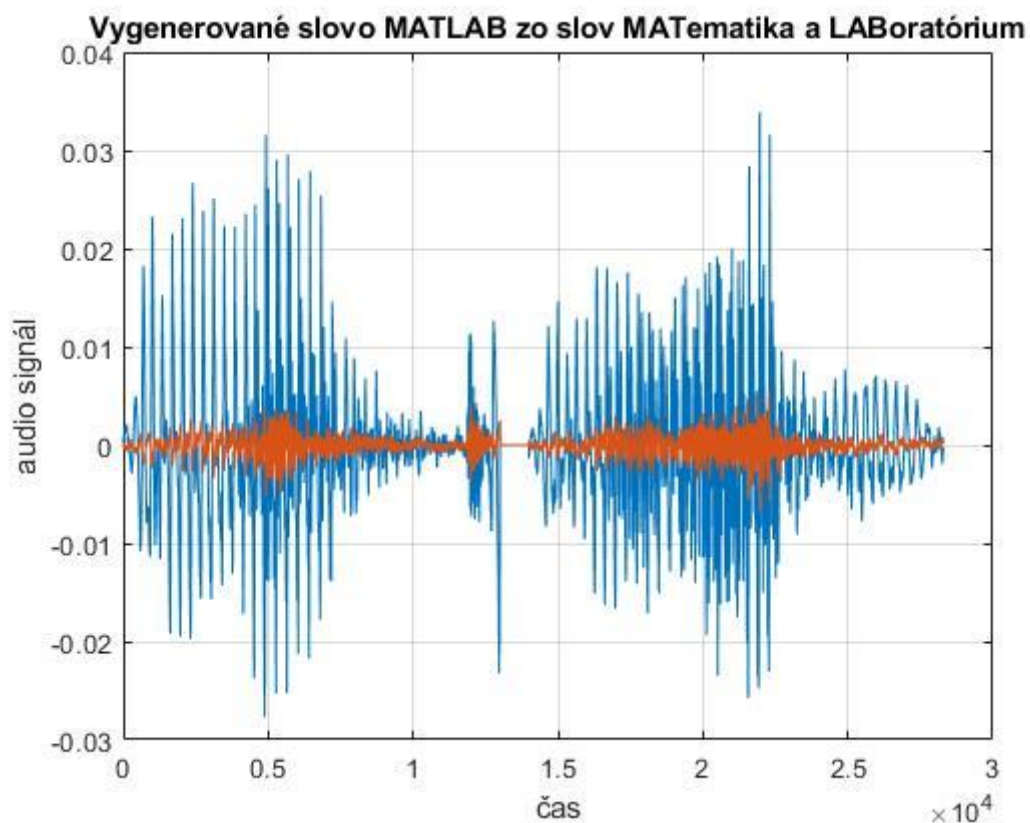
[http://www.fit.vutbr.cz/study/courses/ZRE/public/pred/12_synteza_matousek/tts.pdf]

V tomto riešení som nahral slová „matematika“ a „laboratórium“, vyseparoval z nich úvodné slabiky „mat“ a „lab“, zložil ich dokopy s malou prestávkou a porovnal voči nahranému slovu „MATLAB“. Tieto slabiky by sa mali mať rovnaký dôraz ako pri vyslovovaní slova „MATLAB“

Graf pre nahrané slovo „MATLAB“ vyzerá takto:



A graf pre vygenerované slovo touto metódou vyzerá nasledovne:



Grafy sú takmer totožné a vygenerované slovo znie najprirodzenejšie. Ideálnym riešením by bola konkatenačná syntéza, ktorá predpokladá, že reč sa skladá z takýchto („mat“, „lab“) rečových jednotiek. V tejto syntéze sa vytvára reč reťazením rečových segmentov z uloženého inventára. Podmienkou je ale veľké množstvo kvalitných nahrávok.

[http://www.fit.vutbr.cz/study/courses/ZRE/public/pred/12_synteza_matousek/tts.pdf – str.43]