

Synteza mowy

Ćwiczenia 2

Zajęcia 6

Zadanie 1. (1p) Rozważamy syntezytor sylabiczny. Na wykładzie mówiliśmy o następującym sposobie wyboru bazy sylab: dla pewnego zbioru fraz wyznaczamy ich fonetyczną reprezentację, dzielimy na sylaby, a następnie zliczamy liczby wystąpień poszczególnych sylab i wybieramy K najczęstszych, które będą stanowić naszą bazę.

Przyjmijmy, że naszym celem jest wybór K sylab w ten sposób, by zmaksymalizować liczbę fraz (z zadanego zbioru), które zostaną wypowiedziane bezbłędnie (faza jest wypowiedziana bezbłędnie, jeżeli wszystkie jej sylaby znajdują się w bazie). Czy metoda opisana w tym zadaniu jest dobrym sposobem maksymalizacji liczby bezbłędnie przeczytanych fraz? (czy może ona być przypewnych założeniach optymalna? czy realne jest żądanie optymalności bez dodatkowych ograniczeń? W jakich sytuacjach ta metoda działa źle?)

Zadanie 2. (1p) Zaproponuj inną metodę wyboru sylab do bazy syntezytora sylabicznego.

Zadanie 3. (1p) Nasz syntezytor sylabiczny nie mając sylaby w bazie powinien wybrać jakąś inną, możliwie podobną. Zaproponuj jakąś miarę podobieństwa sylab (to znaczy przypisania każdej parze sylab liczby rzeczywistej w ten sposób, by dla sylab podobnych była ona niewielka). Wyobraź sobie następnie działanie syntezytora, który korzysta ze zbioru sylab podanego na KNO (300 sylab, czyli ten mniejszy) i wybiera na najbliższą sylabę wówczas, gdy nie ma dokładnego odpowiednika w bazie. Jak wymówiłby on Twoje imię i nazwisko?

Zadanie 4. (2p) To zadanie jest w większości powtórką z wykładu (a 2 punkty są tak ogólnie na zachętę). Syntezytor difoniczny to taki, w którym mowę syntetyzujemy ze sklejanых difonów, czyli par głosek. Oprócz sklejanя głos przekształcamy, żeby dać akcenty, intonację lub zmienić tempo mowy.

- Zaprojektuj format bazy difonów (innymi słowy masz podać, jakie informacje dla każdego difonu konieczne są do przeprowadzenia syntezy).
- Teoretyczna liczba difonów jest kwadratem liczby głosek z języka. Podaj przykłady dla języka polskiego świadczące o tym, że nie wszystkie pary głosek są możliwymi difonami (a zatem wspomniane oszacowanie jest zbyt pesymistyczne)
- Przedstaw skrótowo działanie algorytmu syntezy, koncentrując się na p o dobieństwach i różnicach z syntezytorem unifonicznym.
- Dlaczego w bazie difonów warto umieszczać difony, których lewą lub prawą stronę jest cisza?

Zadanie 5. (1p) Zajmiemy się przez chwilę akcentem zdaniowym, czyli silniejszym wypowiedaniem słów w zdaniu. Pokaż na przykładach, że syntezytor który akcent zdaniowy wyznacza w pełni zgodnie z oczekiwaniami użytkownika, powinien posiadać bardzo duże zdolności „rozumienia i interpretacji” wypowiedzanego tekstu.

Przykłady (co najmniej 2) powinny być skonstruowane następująco:

powinieneś znaleźć zdanie, w którym naturalne jest zaakcentowanie dwóch różnych słów, przy czym to, które powinno być zaakcentowane zależy bardzo silnie od poprzedzające to zdanie kontekstu). Powinieneś podać to zdanie i dwa konteksty.

Zadanie 6. (1p) Zaproponuj jakąś prostą metodę wyznaczania akcentu zdaniowego, pokaż na przykładach, jak ona działa (to tylko wprowadzenie do tematu, będziemy o tym jeszcze mówić na wykładzie).

Zadanie 7. (1p) Wyobraźmy sobie, że mamy bazę do syntezytora difonicznego, w której zaznaczone są poszczególne epoki oraz miejsca gdzie zaczyna i kończy się przejście między jedną głoską a drugą.

Przyjmijmy, że nie przejmujemy się zmianą F_0 , a jedynie czasem trwania dźwięku. Wyjaśnij, dlaczego w takiej sytuacji mamy dużo możliwości sklejanя dwóch difonów. Zaproponuj jakąś metodę wyboru jednej z tych możliwości, która daje szansę na lepszy wynik niż algorytm, który wybiera losowy koniec epoki poza obszarem przejścia jako punkt sklejanя (i losowe epoki do replikacji, ewentualnie wycięcia).

Zadanie 8. (1p) Załóżmy, że dysponujemy wielką bazą wszystkich nagrań p. Krystyny Czubówny (czyli lektorki, która mówi wszystko bardzo poprawnie, wyraźnie i w stałym tempie) wraz z transkrypcją fonetyczną i dokładną informacją o czasach fonemów.

Jak wykorzystać te dane do wyznaczania czasów trwania poszczególnych głosek w syntetyzowanej wypowiedzi? Uwaga: interesuje nas metoda, która możliwie jak najsilniej wykorzysta wielkość bazy i możliwie rzadko wykorzystuje pojęcie „średniego czasu fonemu, który...”.

Zadanie 9. (1p)★ Znajdź informację o wybranym dialekcie języka polskiego (lub silnie powiązanym z polskim języku, takim jak śląski czy kaszubski). Przedstaw informacje istotne dla osoby, która chce tworzyć syntezytor mówiący w tym języku (dialekcie). Zakładamy przy tym, że osoba ta potrafi utworzyć syntezytor dla języka polskiego, więc interesują ją jedynie różnice.