



NISLab document



NISLab

**Laboratoriet for Naturligt Interaktive
Systemer, Syddansk Universitet, Odense**

Evaluering af Taletræner for Dansk

*Første Evaluering baseret på Felldata indhentet fra Træningssteder December
2004 – primo Maj 2005*

31. Maj 2005

Forfattere

*Niels Ole Bernsen¹, Thomas K. Hansen¹, Svend Küllerich¹ og Torben Kruchov
Madsen¹*

1: NISLab, Odense, Denmark

Projekt	Udtaletræner for Dansk
Deliverable status	Offentlig
Afleveringsdato	31. Maj 2005
Titel	Evaluering af Taletræner for Dansk. Første Evaluering baseret på Feltdata Indhentet fra Træningssteder December 2004 - Maj 2005
Deliverable art	Rapport
Status & version	Endelig
Antal sider	32
Ansvarlig partner	Laboratoriet for Naturligt Interaktive Systemer (NISLab), Syddansk Universitet, Odense, Danmark
Editor	Niels Ole Bernsen
Author(s)	Niels Ole Bernsen, Thomas K. Hansen, Svend Kiilerich, Torben Kruclov Madsen, alle NISLab
Modtager	Flygtninge- og Indvanderministeriet som vi gerne vil takke varmt for deres støtte til det danske udtaletrænerprojekt
Nøgleord	Automatisk taletræning, dansk udtale, nydanskere
Abstrakt	Denne rapport beskriver resultaterne af den først større felttest med en automatisk udtaletræner som hjælper nydanskere med at forbedre deres udtale af enkeltord på dansk.

Table of Contents

1	Resumé og oversigt	1
1.1	Resultaterne	1
1.2	Plan for rapporten.....	2
2	Testforløbet	3
2.1	Træningssteder	3
2.2	Generelle erfaringer.....	5
3	Metodologi	8
3.1	Indsamling af testfiler.....	8
3.2	Sortering af filer fra træningsstederne.....	8
3.3	Udvælgelse af testfiler til evaluering	8
3.4	Primær evalueringsalgoritme	9
3.5	Træningsmængde	10
3.6	Sekundær evaluering	10
3.7	Trænings- og testbetingelser	11
3.8	Usikkerheder ved evalueringen	12
4	Grafer og eksempler	13
4.1	Studieskolen	14
4.2	Monitor.....	14
4.3	FO.....	15
4.4	Riisingprojektet	16
5	Analyse	17
5.1	Overblik over resultaterne	17
5.1.1	Forklaringer til Tabel 5.1	19
5.1.2	Kommentarer til Tabel 5.1	19
5.1.3	Anomalierne i Tabel 5.1	20
5.2	Træningsmængde og resultater	20
5.2.1	Træningsprisen for 10% fremgang.....	21
5.2.2	Lineært fremskridt?	21
5.2.3	Konklusion om fremskridt som funktion af træningsmængde.....	22
5.3	Individuelle forskelle og forskelle i modersmål.....	22
5.4	Elever med samme testbetingelser	24
5.5	Kontrolmålinger	25
5.6	Træningsbetingelser - igen	26
5.7	Anbefalinger.....	26
5.7.1	Brugermanual og elevinstruktion på træningsstedet	26
5.7.2	Taletrænersystemet.....	27
6	Konklusioner	28

1 Resumé og oversigt

Denne rapport fra den afsluttede anden fase i det danske Udtaletrænerprojekt beskriver resultaterne af den første felttest af taletræneren udført på en række skoler og andre institutioner hvor der undervises i dansk for nydanskere. Der fokuseres på resultaterne af elevernes brug af den automatiske talegenkender til træning af udtale af enkeltord på dansk. Systemet kaldes i det følgende for *taletræneren*. Systemet er udviklet af NISLab som også har stået for rådgivning til institutionerne om installation og brug af systemet, tilsendelse af logfiler fra elevernes brug af systemet til NISLab, mv.

1.1 Resultaterne

Den første felttest med taletræneren har givet et stort (+800 logfiler) og særdeles værdifuldt materiale om taletrænerens brug i praksis på træningsstederne samt om taletrænerens reelle nytteværdi til at forbedre dansk udtale af enkeltord hos personer med et andet modersmål end dansk. Ialt 4 skoler i Odense-området har bidraget det datamateriale vedrørende ialt 22 elever som analyseres i denne rapport. Andre 5 skoler som har modtaget taletræneren til installation, alle uden for Odense-området, har enten ikke bidraget med data overhovedet eller har bidraget med data der var for sparsomme til at kunne bruges i analysen.

Vi kan konstatere, at stort set alle de elever på træningsstederne hvis fremskridt analyseres i denne rapport, har gennemført seriøs og målbevidst træning. Det viser, at taletræneren faktisk virker i praksis, og gør det sandsynligt at taletræneren understøtter elevernes motivation til at træne med systemet.

Vi kan også konstatere, at eleverne som helhed stadig har trænet for lidt til at give os det ideelt ønskede udbytte af evalueringen af taletrænerens brug og nytte på træningsstederne. Eleverne har i gennemsnit gennemført godt 1/4 af det fulde træningspensum på 4500 ord. Da stort set alle træningssteder var i gang pr. 1.2.2005 repræsenterer vores datamateriale mindst ca. 12 ugers data fra de leveringsdygtige træningssteder. Mange elever stoppede dog deres træning væsentligt tidligere end efter 12 uger. Da vi ikke har direkte adgang til eleverne på træningsstederne må vi fremover finde på andre midler til at forsøge at intensivere elevernes træning. Det skal dog i denne forbindelse siges, at felttræningen med taletræneren fortsætter i Odense-området frem til sommeren 2005, så vi håber til den tid at have fået det ønskede materiale om (næsten) fuldt gennemførte træningsperioder for i det mindste nogle elever.

Hidtil er taletrænerens brug, brugbarhed, og nytte kun analyseret i laboratorieomgivelser hos NISLab, bl.a. med grupper af kinesiske og finske studenter, baselining med indfødte danskere, mv., men tests i laboratorieomgivelser er noget helt andet end test i felten, og kun sidstnævnte kan give realistisk information om systemets virkelige anvendbarhed.

Det er efter den foretagne felttest klart at eleverne gør gode fremskridt med taletræneren. I gennemsnit har de 16 elever i hovedgruppen som ikke allerede taler et godt dansk, opnået et fremskridt, som minimum, på 16.4% i absolutte procenttal over deres startscore, og med en gennemsnitlig træningsindsats på ca. 20% af det fulde anbefalede træningspensum. Dette gennemsnit dækker over meget store individuelle forskelle i elevernes fremskridt, hvilket bl.a. betyder at mange elever har gjort væsentligt større fremskridt end gennemsnittet viser. Vi vil anse det for et konservativt skøn at elevernes gennemsnitlige fremgang ved brug af taletræneren med en væsentligt større træningsindsats, altså en indsats som nærmer sig det fulde anbefalede pensum, vil kunne nå mindst 40% absolutte procent over deres starttestscore, således at en elev som i starttesten opnår en score på 30% i sluttesten vil opnå en score på 70% under de samme testbetingelser som i starttesten.

Fænomenet *testbetingelser* er en opdagelse vi har gjort i forbindelse med den her beskrevne felttest og som diskuteres udførligt i rapporten, inklusive de forholdsregler i form af brugermanual- og systemrevision vi uddrager af fænomenet.

Vi mener, at de her fremlagte resultater giver et ret overbevisende grundlag for at antage, at taletræneren er et nyttigt redskab til selvtræning i dansk udtale for personer med et andet modersmål end dansk. De forbehold, vi er nødt til at tage, er at vi gerne vil se resultaterne fra et passende antal elever som har gennemført det fulde træningspensum, og at vi gerne vil se på en række sammenhænge mellem elevernes fremskridt under træningen og deres personlige data, herunder også data om deres brug af lyttetræneren.

Med disse forbehold mener vi at denne rapport etablerer et forsvarligt grundlag for at fortsætte udtaletrænerprojektet med henblik på dansk taletræning på sætningsniveau.

Efter projektets plan for dets anden fase skulle nærværende rapport have været en af fire rapporter som tilsammen gav et fuldstændigt billede af den udførte felttest. De tre andre planlagte rapporter er Myram og Gregersen: Erfaringer på træningsstederne; Gregersen: Evaluering af lyttetræneren; og Gregersen: Personlige data om forsøgspersonerne. Disse rapporter foreligger ikke, hvilket har følgende konsekvenser for nærværende rapport: (1) vi kan ikke sammenholde de modtagne data fra træningsstederne med forholdene på træningsstederne eller integrere anbefalingerne i denne rapport med anbefalinger om træningens organisering på træningsstederne med henblik på at give eleverne det bedst mulige udbytte af den installerede taletræner; (2) vi kan ikke sammenholde elevernes fremgang med brug af lyttetræneren med deres fremgang med brug af taletræneren; og (3) vi kan ikke sammenholde elevernes individuelle fremgang med deres personlige data. Sidstnævnte (3) vil i princippet kunne give særdeles nyttige oplysninger om sammenhænge mellem, f.eks., elevernes træningsfremgang og faktorer såsom modersmål, alder, opholdstid i Danmark, sprogscoleopholdstid, brug af lyttetræneren i forbindelse med brug af taletræneren, osv. Vores data viser f.eks. at den elev som har trænet mest af alle ikke har gjort nævneværdige fremskridt i dansk udtale. Det ville være værdifuldt at forstå hvorfor, men en fuld forståelse kræver muligvis både personlige data om eleven og data om forholdene omkring elevens træning.

1.2 Plan for rapporten

I de følgende afsnit beskrives først felttestforløbet, inklusive data om træningssteder, elever, og modtagne logfiler, samt nogle generelle erfaringer og hvad vi på grund af disse agter at gøre anderledes fremover (Afsnit 2). Derpå præsenteres den metodologi vi har anvendt i sortering og analyse af logfilerne fra elevernes træning og selvtests, og vi nævner nogle af de mange uafhængige variable som er involveret i analysen af feltresultaterne, mange af hvilke vi ved for lidt om - i det mindste indtil andre resultater af Udtaletrænerprojektet foreligger (Afsnit 3). I Afsnit 4 illustrerer vi en del af de målinger der er foretaget pr. elev baseret på logfilerne, og vi viser fremskridtsgrafer for alle 22 elever organiseret pr. træningssted. Afsnit 5 er analyseafsnittet, hvor vi dels præsenterer flere måleresultater pr. elev, dels diskuterer datamaterialet som helhed fra en række perspektiver. Afsnit 6 konkluderer rapporten.

2 Testforløbet

Denne rapport fra det danske udtaletrænerprojekt beskriver resultaterne af den første felttest af taletræneren udført på en række skoler og andre træningssteder hvor der undervises i dansk for nydanskere. Udtaletræneren består af en lytte-komponent hvor eleverne kan afprøve deres evne til at skelne mellem minimalt forskellige danske ordpar, og en taletrænerkomponent hvortil eleverne kan udtale enkeltord og få umiddelbart feedback på hvor godt ordene er genkendt af taletrænerens talegenkender. I det følgende beskrives brugen af taletræneren i de udførte felttests samt de opnåede resultater.

2.1 Træningssteder

Der har været stor interesse for taletræneren. Tabel 2.1 viser at taletræneren er sendt til installation på 9 træningssteder. Installation på det første træningssted fandt sted i november 2004. Siden installeredes i december 2004 og januar 2005, hvorefter de installationer vi har kendskab til stort set var gjort. Det vides ikke om alle de anførte træningssteder faktisk har installeret taletræneren når de har modtaget den, eller hvad der er gået galt hvis de ikke har installeret systemet. Alle træningssteder har fået information om at de blot skal kontakte NISLab hvis de har problemer med installation, brug, eller fremsendelse af logfiler.

Træningssted	Geografi	Antal logfiler fra træningssted	Antal elever i træning	Antal elever med målbare resultater
Aalborg Sprogcenter	Aalborg	0	0	0
AOF	Svendborg	19	7	0
EUC Nord	Hjørring	0	0	0
FO	Odense	228	25	4
Monitor	Odense	252	15	10
Nordjyllands Idrætshøjskole	Aalborg	0	0	0
Riisingprojektet	Odense	91	7	4
Studieskolen	Odense	203	19	4
S&S	Svendborg	28	15	0
Ialt	9	821	88	22

Tabel 2.1. Oversigt over træningssteder, logfiler og elever.

Projektets formål med installation på en række træningssteder har været dobbelt. Dels har det været vigtigt (1) at indhente erfaring med installation og brug af systemet på forskellige træningssteder. En analyse af disse erfaringer er essentiel for at kunne optimere installation og brug af systemet i fremtiden. Træningsstederne er nemlig meget forskellige mht., blandt andet:

- elevpopulation
- teknisk udstyr
- teknisk ekspertise
- ledelse og formål

- personalets mulighed for, og motivation til, at dedikere den fornødne tid til installation, kommunikation med NISLab, og motivation af stedets elever til at bruge systemet

Derfor er det vigtigt med henblik på senere massiv installation af systemet at indhente erfaringer med brugen af systemet under hvad der i det foretagne felteksperiment har vist sig at være vidt forskellige betingelser. Erfaringsindhentningen bør også omfatte effektiviteten af den givne tekniske støtte til træningsstederne, evt. mangler ved den første udgave af brugs- og brugermanualen til systemet, mv.

Desuden (2) har formålet med installation af systemet på træningsstederne naturligvis været at indhente resultater som kan belyse i hvilket omfang eleverne faktisk gør fremskridt gennem taletræning med systemet.

NISLab har i felttestperioden ydet støtte til installation af systemet på træningsstederne, dels i form af fremmøde på træningssteder i Odense-området, dels i form af email og telefonisk støtte til alle træningssteder. Ud over denne instruktion har træningsstederne modtaget den første udgave af brugs- og brugermanualen til systemet.

Som det fremgår af Tabel 2.1 har tre træningssteder ikke leveret en eneste logfil. To andre træningssteder har ganske vist leveret logfiler, men intet sæt af disse har opfyldt kriterierne, beskrevet i Afsnit 3 nedenfor, for at være brugbare til evaluering af elevernes fremskridt. Blandt de mulige grunde til den manglende frembringelse af brugbare testresultater kan nævnes:

- man mistede modet da man modtog systemet, f.eks. af tidsmæssige grunde for personalet på stedet
- man havde tekniske vanskeligheder med at få systemet til at køre på træningsstedet som ikke kunne overvindes, evt. trods anmodning om NISLabs støtte
- man havde tekniske vanskeligheder med at fremsende logfiler som ikke kunne overvindes, evt. trods anmodning om NISLabs støtte
- man producerede for få logfiler til at tillade os på NISLab at bedømme fremskridtene hos eleverne på træningsstedet, dvs. at der var for lidt trænings- og testaktivitet på træningsstedet

Til gengæld har 4 træningssteder leveret et stort antal logfiler som til sammen har tilladt os at analysere fremskridt ved brug af taletræneren for 22 elever ud af de ialt 88 elever som har forsøgt sig med systemet.

Det synes klart at et stort antal elever kun har ”snuset” til taletræneren uden at bruge den mere intensivt til at forbedre deres danske udtale. Dette kan skyldes mange faktorer. F.eks. fremgår det af Tabel 2.1 at FO, Monitor og Studieskolen har leveret stort set lige store mængder logfiler til NISLab. Ikke desto mindre har kun 4 elever ud af 25 på FO, og 4 elever ud af 19 på Studieskolen, leveret tilstrækkeligt med data til analyseformål, medens hele 10 elever ud af 15 på Monitor har leveret tilstrækkeligt med data. Disse forskelle er så store, at der kunne ligge noget bag som det er vigtigt at kende til med henblik på at sikre maksimal succes i den fremtidige brug af taletræneren i felten. En mulig forklaring kunne være at personalet på Monitor har taget sig væsentligt mere af eleverne, opfordret dem til at træne med taletræneren, stået dem bi i forbindelse med træningen, og generelt udvist større entusiasme over for eleverne. Riisingprojektet har leveret langt færre logfiler end de tre allerede nævnte skoler. Ikke desto mindre har vi analyserbare data for 4 ud af 7 elever fra Riisingprojektet, hvilket er sammenligneligt med resultaterne fra Monitor. Det modsatte ekstrem, så at sige, ses f.eks. på S&S hvor hele 15 elever kun har leveret 28 logfiler tilsammen. Hvis disse store forskelle ikke primært skyldes personalets holdning og støtte, er en anden forklaringsmulighed at elevpopulationerne er meget forskellige på træningssteder, som udviser store

forskelle i forholdet mellem antallet af elever som har prøvet taletræneren og antallet af elever som har gennemført substantiel træning med systemet.

På i det mindste tre af de 4 aktive træningssteder fortsætter udtaletræningen i skrivende stund med et betydeligt antal elever. Tabel 2.1 viser at det faktisk kun var de 4 træningssteder i Odense som leverede brugbare logfilesæt. Dette kan meget vel skyldes at disse træningssteder modtog instruktion fra NISLab gennem personligt fremmøde på træningsstedet, i nogle tilfælde mere end 10 gange. Men vi har alligevel ingen god forklaring på hvorfor antallet af modtagne logfiler falder proportionalt med et træningssteds geografiske afstand fra NISLab.

For alle træningssteder gælder det, at alle elevers trænings- og testsessioner med systemer logges og logfilerne sendes til NISLab til sortering og, hvis logfilerne for en elev opfylder projektets kriterier for målbarhed, efterfølgende analyse.

2.2 Generelle erfaringer

Såvidt vides har træningsstederne generelt modtaget taletræneren med stor entusiasme – i det mindste umiddelbart. Samtidig har det vist sig, at flere træningssteder af mange forskellige grunde ikke har haft eller kunnet afse den fornødne tid og ekspertise til at skabe evaluerbare resultater med taletræneren. Dette var forventet fra starten, og blev foregrebet ved at rekruttere et større antal træningssteder end nødvendigt fra en synsvinkel som alene fokuserer på testresultater og deres analyse.

Der er indkommet et stort antal logfiler fra træningsstederne, 821 i alt, og de viser generelt at nogle aspekter af installation og brug af taletræneren bør forbedres i de kommende måneder. Systemet selv er i orden, både teknisk set og mht. dets brugbarhed til praktisk selvtræning, så nogle træningssteders vanskeligheder skyldes andre faktorer. Det samme gælder fremsendelsen af logfiler til NISLab.

Det største problem som skal løses angår elevernes træning med systemet. Eleverne kan både træne og teste sig selv med umiddelbart feedback på deres præstation. Men (1) systemet skelner pt. ikke mellem trænings- og testsituationer. Det betyder at logfilerne skal adskilles manuelt på NISLab før evaluering kan ske. Desuden (2) er brugermanualen ikke håndfast nok i sine instruktioner til eleverne og deres evt. lærere på træningsstederne med hensyn til at anbefale hvordan de skal træne og teste. Det betyder, at sammenlignelige testresultater fra en bestemt elev ikke genereres med den ønskede regelmæssighed.

Disse to problemer er nu ved at blive løst, (1) ved at forsyne systemet med funktionalitet så eleverne selv kan vælge om de vil træne eller teste. Testfilerne markeres så som testfiler, og det er disse som den efterfølgende evaluering vil koncentrere sig om, medens træningslogfilerne kun tages i betragtning som data der belyser den pågældende elevs samlede træningsindsats. Desuden betyder dette, at eleven straks efter en test kan få feedback på det samlede testresultat. (2) løses ved en ny version af brugermanualen som kraftigt anbefaler, for ikke at sige, dikterer, en bestemt fremgangsmåde i træning og test med henblik på at lette den efterfølgende evaluering. Det har vanskeliggjort resultatanalysen at eleverne ikke med en passende regelmæssighed forsøger at udtale de samme sekvenser af systemets ialt 450 danske ord.

Det bør i denne forbindelse understreges, at vi står over for at skulle skabe en helt ny ”kultur” i brugen af ny informationsteknologi til selvtræningsformål. Ingen er bekendt med, endsige vant til, denne kultur, og felttestkarakteren af den her beskrevne taletrænerevaluering har betydet, at vi ikke har været i stand til at kommunikere direkte med eleverne for at ændre deres testadfærd. Vejen til eleverne går gennem træningssted, personer på træningsstedet som måske, måske ikke, selv instruerer eleverne, og en ny version af brugermanualen som man i projektet ikke har skønnet skulle udsendes midt i felttestperioden. Derfor er vi nødt til at vente med de igangværende forbedringer indtil en ny felttestperiode gennemføres.

Et andet vigtigt punkt er, at analysen af de hidtidige testresultater nedenfor har vist et for os helt nyt fænomen som den næste version af brugermanualen, og muligvis også af systemet, skal tage hensyn til, nemlig de mere præcise *testbetingelser* som eleverne tester under. Denne opdagelse diskuteres udførligt i nærværende rapport.

Et helt centralt fænomen som er afsløret af de modtagne testfiler er, at stort set alle testpersoner endnu ikke har trænet og testet længe nok. De første træningssteder modtog taletræneren i november-december 2004 og de fleste af dem havde taletræneren kørende før udgangen af januar 2005. Brugermanualen understreger klart nok at træning er afgørende for at lære dansk udtale, og man kan sagtens argumentere for at den træningsmængde der i denne rapport er sat som vejledende, dvs. udtale af 4500 ord i løbet af 10 uger, er for lille og burde forøges. Men dels som følge af nødvendigheden af at rapportere resultater på nuværende tidspunkt, og dels som følge af den de facto gennemsnitligt for lille træningsmængde per elev pr. uge som vi kan konstatere i de modtagne resultater, er det vanskeligt i denne rapport at drage den konklusion vi allerhelst ville kunne præsentere. Denne er nemlig at kunne *bevise* fra analysen af de modtagne data, at med en ”fuld” træningsmængde kunne et overbevisende antal af elever udtale enkeltord på dansk næsten som danskfødte kan gøre det, dvs. med en testscore på +70% og under de vanskeligste testbetingelser.

Hvad vi kan demonstrere nedenfor er at forsøgspersonerne generelt har opnået gode fremskridt med taletræneren.

Tabel 2.2 viser et eksempel på en logfil. Første kolonne giver ordets nummer; anden kolonne giver ordets skrevne form, inklusive, i uregelmæssige tilfælde, ordets simplificerede lydskrift; tredje kolonne viser hvor mange gange testpersonen har set ordets ortografi; fjerde kolonne viser hvor mange gange testpersonen har hørt ordet udtalt før personen selv har forsøgt at udtale det; femte kolonne viser hvor mange gange testpersonen har set og hørt ordet udtalt på video før personen selv har forsøgt at udtale det; og sjette kolonne viser systemets scoring af forsøgspersonens udtale. Det skal bemærkes, at langt fra alle logfiler er lige så regelmæssige som den viste, der kun har et enkelt ”hul” idet ord 10 ikke er forsøgt udtalt. Forsøgspersonen kan f.eks. have undladt at udtale mange eller alle ord, udtalt et ord mange gange i træk, osv. Dette understøtter betydningen af at forsøgspersonerne selv genererer ”rene” testlogfiler gennem en klar separation i systemets grænseflade mellem træning og test.

ID	Ord	Set	Audio	Video	Score
1	Objektiv [objægtiu]	1	3	1	2
2	Fortælle	1	1	1	2
3	Begynde [begøne]	1	1	1	2
4	Føtex [føtæks]	1	1	1	0
5	Arbejde [Abaide]	1	2	1	2
6	Betyde	1	1	2	0
7	Lunge [långe]	1	2	1	2
8	Betale	1	3	1	0
9	Forsøge	1	2	1	0
10	Pile	0	0	0	
11	Fortsætte	1	3	2	0
12	Gade	1	2	3	2
13	Dele	1	1	1	0
14	Opleve	1	2	1	1
15	Billigst [bilist]	1	2	2	0
16	Udvikle	1	3	1	2
17	Bakke	1	2	1	1

18	Cykel	1	2	1	2
19	Nyheder	1	2	1	2
20	Respons [ræspons]	1	2	1	0
21	Brugsen	1	3	2	2
22	Rideklub	1	3	1	0
23	Omhyggelig	1	1	1	0
24	Gymnastik	1	2	2	0
25	Fortryllende	1	2	1	0

Tabel 2.2. Eksempel på logfil.

3 Metodologi

3.1 Indsamling af testfiler

Siden november 2004 er 821 logfiler indsamlet fra 6 træningssteder, jf. Tabel 2.1. Vi har modtaget brugbare testfiler som opfylder kravene beskrevet i de følgende afsnit fra:

- Studieskolen, Odense, 4 elever
- Monitor, Odense, 10 elever
- FO, Odense, 4 elever
- Riisingprojektet, Odense, 4 elever

Det datamateriale der ligger til grund for den følgende analyse udgør således testresultater fra 22 elever fra 4 forskellige træningssteder.

3.2 Sortering af filer fra træningsstederne

Den nuværende version af taletræner + brugermanual opfordrer brugerne til at træne udtale med taletrænerens 450 (nummererede) danske ordforråd og teste sammenhængende serier af ord når de selv vil. Denne kombination eller, om man vil, sammenblanding, af træning og test stiller, har vi konstateret, unødvendige krav til den efterfølgende sortering af logfiler fra træningsstederne. Den næste version af taletræneren vil løse dette sorteringsproblem. Den vil tilbyde eleven at vælge mellem at træne eller teste og markere alle logfiler som enten trænings- eller testfiler. Træningsfilerne vil også blive brugt i evalueringen af brugernes fremskridt, nemlig til at belyse hvor meget og hvor længe de har trænet. Men træningsfilerne vil ikke blive brugt til evaluering af brugerens fremskridt. Dertil bruges kun testfilerne. Disse ting vil naturligvis fremgå af næste version af brugermanualen for taletræneren som samtidig kraftigt vil anbefale at træne og teste på en systematisk og velordnet måde. Derved kan vi forhåbentlig helt undgå tilfælde såsom at elev E starter med at teste ordene 205-267, derpå tester med 55-96, og aldrig gentager en tidligere testsekvens, men højst, f.eks., ”gentester” senere med sekvensen 228-301, hvorved vi kun har fællesmængden 228-267 med den førstnævnte sekvens til rådighed til fremskridtsevaluering.

Pt. er vi nødt til i logfilerne at sortere mellem træningsfiler eller –filfragmenter, og testfiler eller –filfragmenter. Ofte vil brugeren have trænet og testet i samme fil, og træningen behøver ikke at inkludere udtale af danske ord. Den består sommetider i blot at se på de skrevne ord som skal udtales, lytte til deres danske udtale, eller se på og lytte til deres danske udtale på video. Sommetider består en træningsfil i at udtale få ord mange gange for at få dem udtalt rigtigt, osv. Alt dette er naturligvis fint til træningsformål, men disse filer kan ikke bruges til systematisk evaluering af brugerens fremskridt.

Derfor sorterer vi de logfiler fra som kun eller overvejende kun består i træning. Tilbage bliver en delmængde af alle logfiler fra en bestemt elev som indeholder testudtale af ord i rækkefølge. Vi kan kalde disse filer for ’testfiler’.

3.3 Udvalgelse af testfiler til evaluering

Testfilerne organiseres i mapper pr. træningssted og pr. elev på træningsstedet. De organiseres dernæst efter testdato og efter hvilke sekvenser af ord som udtaltes på datoen. Dette indskrives i logfilekstensionen. Derpå udvælges testfiler til evaluering. Det sker ved først at udvælge, for hver enkelt elev, testfiler hvori brugeren (1) tester med *samme* fragment af de 450 testord i rækkefølge og (2) gør dette med et betydeligt tidsinterval til forskel. (1) betyder f.eks., som illustreret ovenfor, at fragmenterne 375-431 og 400-450 kun har ordene 400-431

til fælles. (2) betyder f.eks. at elev E's test med ordene 150-200 d. 20.12.2004 og med de samme ord d. 2.1.2005 er værdiløs til evalueringsformål, fordi eleven ikke har trænet overhovedet imellem de to tests.

I tilgangen til analysen af testfilerne har det for os på NISLab været et empirisk spørgsmål i hvilket omfang betingelserne (1) og (2) ovenfor har været opfyldt i logfilerne. Givet den relativt lille træningsmængde i gennemsnit pr. elev i datamaterialet, har det for det meste kun været muligt at finde +25 ords fragmenter gentaget i datamaterialet 2 gange. Sommetider er fragmentet gentaget 3 gange, i hvilke tilfælde vi også har foretaget en måling på det mellemste fragment i tid. I gennemsnit for alle elever er tidsintervallet mellem to tests brugt til evaluering = 5.3 uger. I to tilfælde er vi gået ned til 1 uges tidsinterval, dvs. op til 10 dage mellem første og sidste test. Resultaterne af disse korte trænings- og testintervaller er naturligvis behæftet med en større usikkerhed end resultater baseret på 5-10 ugers træning og med en tilsvarende større træningsmængde. Antallet af ord i sekventielt ordnede fragmenter som sammenlignes er mindst 25 identiske ord. Vi ville have foretrukket min. 50 ord men det har ikke været muligt.

3.4 Primær evaluering algoritme

Da de fleste brugere følger brugermanualens vejledning om at forsøge at udtale alle 450 ord i rækkefølge over en periode, kan vi definere begrebet om en 'testcyklus'. En testcyklus er et tidsinterval i hvilket en bruger udtaler alle 450 ord mindst en gang. Dette tager som regel ganske lang tid, såsom 6 uger, også fordi mange brugere går langsomt frem og desuden træner med de samme, f.eks. 100 ord, flere gange før de går videre til næste ordsekvens. Som bemærket i Afsnit 3.3 kan vi ikke bruge i tid helt nærtliggende, gentagne testsekvenser til evalueringsformål. Da det første træningssted begyndte at levere data i november 2004 og mange af de øvrige først kom i gang sent i januar 2005 eller først i februar 2005, kan man regne ud at kun relativt få brugere har genereret mere end to testcykler indenfor hvilke identiske testsekvenser er tilstrækkeligt adskilt i tid til at evaluering er meningsfuld.

Metoden til at udvælge testfiler til evaluering kan derfor sammenfattes som følger:

- find bruger X på træningssted Y som mindst har gennemført +1 testcyklus. Det er nok at brugeren har testet alle 450 ord + startet på næste testcyklus;
- find en sekvens S på mindst 25 ord (og helst flere) som X har udtalt mindst 2 gange;
- foretræk at S er udtalt med mindst en gennemsnitlig testcyklus' mellemrum i tid, dvs. ca. 6 uger;
- rens ud i de udvalgte logfiler ved at fjerne gentagelser af et ords udtale, noter evt. "huller" hvor ord i ordsekvensen ikke er udtalt, mv. Derved undgås fejlbehæftet statistik;
- beregn i % korrektheden af X's udtale af S første gang, n'te gang og sidste gang den optræder i det foreliggende materiale;
- korrekthed beregnes ved at addere elevens score på testsekvensen af udtalte ord og beregne scoren i % af maksimalscoren, dvs. en score på 2 pr. testord;
- plot X's fremskridt eller tilbageskridt i udtalekorrekthed i en graf for det pågældende træningssted;
- gør dette for alle testpersoner på træningssted Y og fortsæt til træningssted Z;
- gentag algoritmen for træningssted Z.

Resultatet af denne procedure er, for X, Y, og S, en graf for tidlig, evt. mellemliggende, og sen testning.

3.5 Træningsmængde

Brug af algoritmen i afsnit 3.4 som plottet i testgrafene tager ganske vist højde for den *tid* der ligger imellem startresultatet Rst, mellemresultatet Rml, og slutresultatet Rsl, men algoritmen tager ikke højde for den *træningsmængde* elev X har investeret i at komme fra startresultatet Rst til slutresultatet Rsl. Vi er også interesseret i at vide noget om effekten af brug af taletræneren for forskellige elever. Måske ”virker den ikke” for nogle elever, medens den ”virker strålende” på andre. Det kan man ikke se af at, f.eks., både elev E1 og elev E2 er gået fra startresultat 50% til slutresultatet 60%. Hvis E1 har klaret dette på nogle få uger og med beskedent træning, medens E2 har klaret det på 13 uger med massiv træning, så fortæller dette os et eller andet om forskelle mellem elever, såsom om deres individuelle lærenemhed, deres modersmåls indflydelse på lærenemheden, osv., og taletrænerens effektivitet i den forbindelse. Hvad det præcist fortæller os kan imidlertid kun afgøres hvis vi har personlige data om hver elev som vi evaluerer, og det har vi som sagt ikke for øjeblikket.

For at tage højde for træningsmængden har vi anoteret alle fremskridtsresultater i Tabel 5.1 med en række parametre som belyser den involverede træningsmængde og som også belyser elevens samlede træningsmængde. Disse annoteringer muliggør vurdering af taletrænerens effektivitet pr. elev og gør det lettere at sammenligne de forskellige elevers resultater. Vi fandt i den forbindelse at træningsmængden kan opgøres på mange forskellige måder, hvoraf nogle er mere vildledende end reelt informative om elevernes faktiske træningsindsats. En af denne rapportes centrale målestokke for elevernes træningsindsats er det antal ord de har udtalt i løbet af træning og testning.

Med henblik på at etablere et mere overordnet, meningsfuldt mål for træningsindsatsen end antallet af trænede ord, har vi defineret begrebet om *et fuldt træningspensum* som = 10 x 450 udtalte ord over 10 uger. Derved kan elevens faktiske træningsmængde angives i % af et fuldt træningspensum. Det er en vigtig diskussion om et fuldt træningspensum faktisk bør være af nøjagtigt det beskrevne omfang. Dette betragter vi som et delvist empirisk spørgsmål, hvis svar må findes ved at se på hvor stor en træningsindsats et gennemsnit af elever skal gennemføre for at nå et passende niveau i dansk udtale. Det er pt. for tidligt at afgøre svaret på dette, eller disse, spørgsmål, eftersom den elevpopulation som analyseres i denne rapport ikke har trænet i tilstrækkeligt stort omfang. Af beslægtede grunde kan vi heller ikke evaluere langtidseffekterne af fremskridt med taletræneren. En sådan evaluering kræver både længere tids træning af et antal elever, en efterfølgende træningspause, en fornyet test, samt formentlig også kendskab til personlige data om eleverne.

3.6 Sekundær evaluering

Man vil aldrig kunne styre elever fuldstændigt når de udfører selvtræning og selvtestning i modsætning til klassetræning og klassetestning. Derfor er det følgende problem en nødvendig følge af elevernes selvstændige arbejde med taletræneren. Som beskrevet ovenfor måles elevernes fremskridt gennem sammenligning af deres udtale af *samme* ordfølgefragment over tid. Samme ordfølgefragment vil kun i usandsynlige tilfælde findes den første dag resp. den sidste dag eleven tester med taletræneren. Derfor har vi også set på et andet mål for træningsfremskridt, nemlig sammenligning mellem tre testfragmenter som *ikke* nødvendigvis alle indeholder de samme ord, men som er karakteriseret ved at de (1) er forskellige fra de i afsnit 3.4 beskrevne testfragmenter, og (2) er aftestet ved start, et tilfældigt valgt sted midt imellem, og til slut i logfilmateriale fra eleven. Dette giver et nyttigt kontrolmål for det i afsnit 3.4 beskrevne mål for elevens fremskridt.

Tilsammen betyder den primære og sekundære evaluering af eleverne, at samtlige elevers fremskridt er målt mindst 5 gange i datamaterialet (logfilerne) for den enkelte elev.

3.7 Trænings- og testbetingelser

I analysen af fremskridtresultaterne i de modtagne logfiler fandt vi somme tider (vha. algoritmen i Afsnit 3.4) et uforklarligt stort *fald* i elevens testpræstation over tid. Sådanne fald kan forekomme to steder i det kvantitative testresultatrum: (1) eleven har et højt startresultat, ofte sammenligneligt med resultatet for indfødte dansktalende, der ligger på 78% og opefter ifølge projektets baseline målinger med indfødte danskere - og næste testresultat viser så et fald på 5-15%. Et sådant *high-drop* betyder ikke noget. Eleven kan, eller er tæt på at kunne, udtale enkeltord på dansk korrekt, og elevens ”dagsform” bestemmer den målte variation som man også finder det hos indfødte danskere.

Vi definerer i forbindelse med denne diskussion *high* som >70% korrekt, *middle* som 50-70% korrekt, og *low* som <50% korrekt.

Derimod var det overraskende at finde *middle-to-low* drops eller endog *high-to-low* drops hos visse elever. Analysen viste her en for os uventet sammenhæng i form af at disse elever havde ændret *testbetingelserne* i start- og sluttesten. Testbetingelserne er om eleven, i testen, før ordene i ordsekvensen udtaltes, havde nøjedes med at *læse* det ord der skulle udtales, hvad alle gjorde; *læse* det og *høre* det på bånd; *læse* det og *se og høre* det på video; eller *læse* det, *høre* det på bånd, og *se og høre* det på video. Hvad vi fandt i datamaterialet var, at de beskrevne fald i præstation synes at kunne forklares ved at eleven havde ændret testbetingelserne *opefter* i følgende hierarki:

- læse
- læse og høre
- læse og se og høre
- læse og høre og se og høre

Vores hypotese er at den ændrede testbetingelse forklarer elevens signifikante drop i udtalepræstation. Med andre ord, alle eller de fleste elever med dansk som andetsprog har lettere ved at præstere korrekt udtale hvis de bruger hele hierarkiet ovenfor som instruktion før de forsøger sig med at udtale et ord end hvis de kun bruger en del af hierarkiet. Denne hypotese er dog for det første ikke fuldt bevist endnu pga. det begrænsede datamateriale, og for det andet er den for upræcis. De indikationer vi har er, at video (*se og høre*) er central for præstationen for de fleste elever. Hvis eleven ikke bruger video før udtale, vælger eleven med andre ord en testbetingelse som er betydeligt sværere end hvis eleven også bruger video-informationen. Derimod kan vi ikke i de foreliggende data se nogen klar effekt af om eleven vælger at bruge eller ikke bruge audio (*høre*). Og vi kan naturligvis heller ikke sige om disse generalisationer gælder for alle elever eller kun for flertallet af dem. Men som vi skal se er evidensen faktisk massiv for at brug eller undladelse af brug af video før udtale er en væsentlig faktor for elevernes testscore.

Testbetingelserne indfører således et nyt sæt uafhængige variable i vores feltdata, variable som medbetingelser elevernes målte fremskridt i udtaletræning. Uden også at se på testbetingelserne for en elevs udtale af en ordsekvens er det formentlig ikke muligt at bedømme i hvilket omfang eleven gør fremskridt. I sig selv er elevens gennemsnitsscore for en udtalesekvens utilstrækkelig.

En anden vigtig konsekvens er, at en evt. anbefaling til den enkelte elev om hvilken yderligere træningsmængde eleven skal regne med for at blive god til dansk udtale af enkeltord, er afhængig af de træningsbetingelser eleven har opnået sine hidtidige resultater under. F.eks. må man konkludere, alt andet lige, at en elev som i træningsforløbet har opnået 57% med taletræneren under brug af både audio og video, er dårligere til dansk udtale end en elev som ligeledes har opnået 57% med taletræneren, men *uden* brug af video. Med et andet eksempel

er en elev som opnår 47% uden brug af video muligvis gennemsnitligt bedre til dansk udtale end en elev som opnår 57% med brug af video.

Sammenligning på lige fod af to elevers præstationer, eller af samme elevs præstationer til på forskellige målepunkter i træningsforløbet, kræver således at de målte testscorer er opnået under de samme testbetingelser. Givet det pt. begrænsede elevmateriale på 22 elever, udgør dette forhold en vis vanskelighed for relevant sammenligning af elevernes præstationer.

3.8 Usikkerheder ved evalueringen

Afsnittene ovenfor har beskrevet den kompleksitet, der er involveret i at evaluere elevernes fremskridt baseret på felttestlogfilerne. Derudover er det vigtigt at gøre klart, at evalueringen i nærværende rapport af elevernes fremskridt er underlagt en række ikke-eliminerbare usikkerheder, såsom:

- den konkrete testsituation: støjkilder, elevens holdning til testningen, osv.;
- tekniske vanskeligheder under testen, f.eks. en mikrofon som svigter, et headsæt som sidder dårligt;
- testinstruktioner på træningsstedet, personlig støtte til eleven, brugermanual;
- brug af lyttetræneren som del af træningen;
- elevens øvrige liv i Danmark, længden af elevens hidtidige ophold i landet, hvor længe eleven har gået på sprogskole, hvor ofte eleven omgås indfødte danskere, osv.;
- elevens modersmål.

Vi kender ikke den konkrete situation i hvilken en elev har testet sig selv med en ordsekvens. Vi antager at eleverne i det store og hele har testet sig selv uden forstyrrelse under testningen. Men vi har intet grundlag for at udelukke, at elevernes selvtestning somme tider er foregået sammen med andre i testrummet, med baggrundsstøj som har forstyrret talegenkendelsen, osv. Noget tilsvarende gælder den enkelte elevs holdning til brug af taletræneren, altså om eleven arbejdet hårdt og målbevidst på at træne og teste eller om eleven ikke har taget det hele særlig højtideligt. Elever som har præsteret relativt større træningsmængder har, går vi på forhånd ud fra, haft en generelt seriøs holdning til træning og testning. Dette bekræftes i høj grad af deres fremskridtsresultater, som vi skal se. Elever som kun har gennemført relativt få trænings- og testsituationer kan vi derimod ikke vide noget helt bestemt om.

Vi kender heller ikke nok til det omfang i hvilket de ansvarlige for brugen af taletræneren på træningsstederne har taget sig af elevernes træning og testning, diskuteret brugermanualen med dem, stået bi under testningen og hjulpet med at beherske teknologien og udstyret, osv. I denne forbindelse ved vi heller ikke om den enkelte elev som vi har logfiler fra har læst brugermanualen eller ej.

Vi ved ikke i hvilket omfang en elevs målte fremskridt i taletræning er påvirket af elevens brug af lyttetræneren parallelt med taletræningen.

Endelig er der naturligvis de vigtige uafhængige variable der har at gøre med elevens øvrige danskindlæring før og under udtalet træningen, elevens sprog- og kulturbaggrund, osv. Dem kender vi ikke.

4 Grafer og eksempler

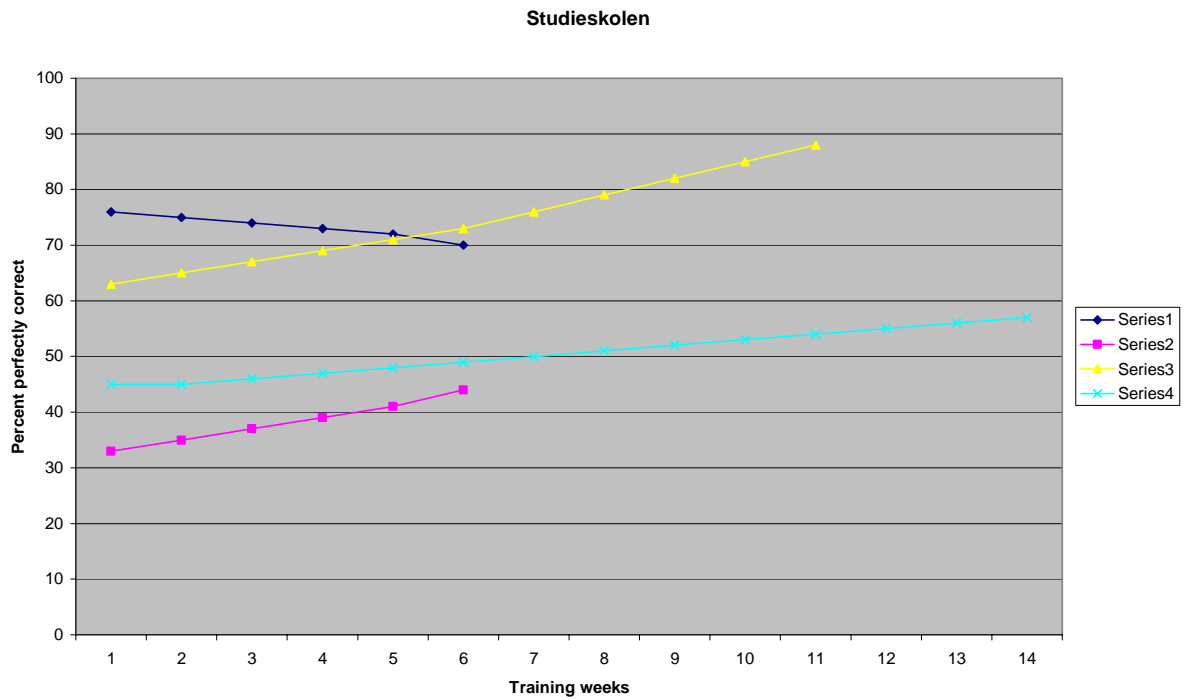
I dette kapitel viser vi fremskridtsgrafer i henhold til den primære evalueringsskema (Afsnit 3.4) for alle 22 elever pr. træningssted.

De i Figur 4.1, 4.2, 4.3, og 4.4 viste grafer er baseret på tabeller indeholdende målinger pr. elev. Et eksempel på en tabeloversigt for en enkelt elev ses i Tabel 4.1. Tabellens højre søjle rummer nogle forklaringer til læseren af denne rapport. Den øverste del af tabellen viser fremskridt i henhold til den primære evalueringsskema, medens den nederste del viser resultaterne af de tre kontrolmålinger pr. elev. Disse diskuteres i Afsnit 5. For målinger af elevernes træningsmængde se Tabel 5.1.

Attribut	S3 = elev id	Forklaringer
Første testdato	1.2.	
Første test antal ord	23	
Ord id	115-139 -2	ordsekvens
Testbetingelse	RA	læse (R) og lytte (A)
Score start	29/46 = 63%	
Score midt	-	ingen mellemscore
Sidste testdato	19.4.	
Sidste test antal ord	25	
Ord id	115-139	ordsekvens
Testbetingelse	RA	læse og lytte
Score slut	44/50 = 88%	
Trend	middle-high	trend verbalt udtrykt
Kommentar	Remarkable progress	
Træningsuger teststart til testslut	10	
Antal logfiler	13	
Scoret sample først eller tidligt	3.2 59/98 = 60%	første sample måling
Ord id	156-204	ordsekvens
Testbetingelse	RA	læse og lytte
Scoret midt-sample	15.2 69/118 = 58%	anden sample måling
Ord id	326-384	ordsekvens
Testbetingelse	RA	læse og lytte
Scoret sample sent eller slut	26.4 66/82 = 80%	tredie sample måling
Ord id	140-180	ordsekvens
Testbetingelse	RA	læse og lytte

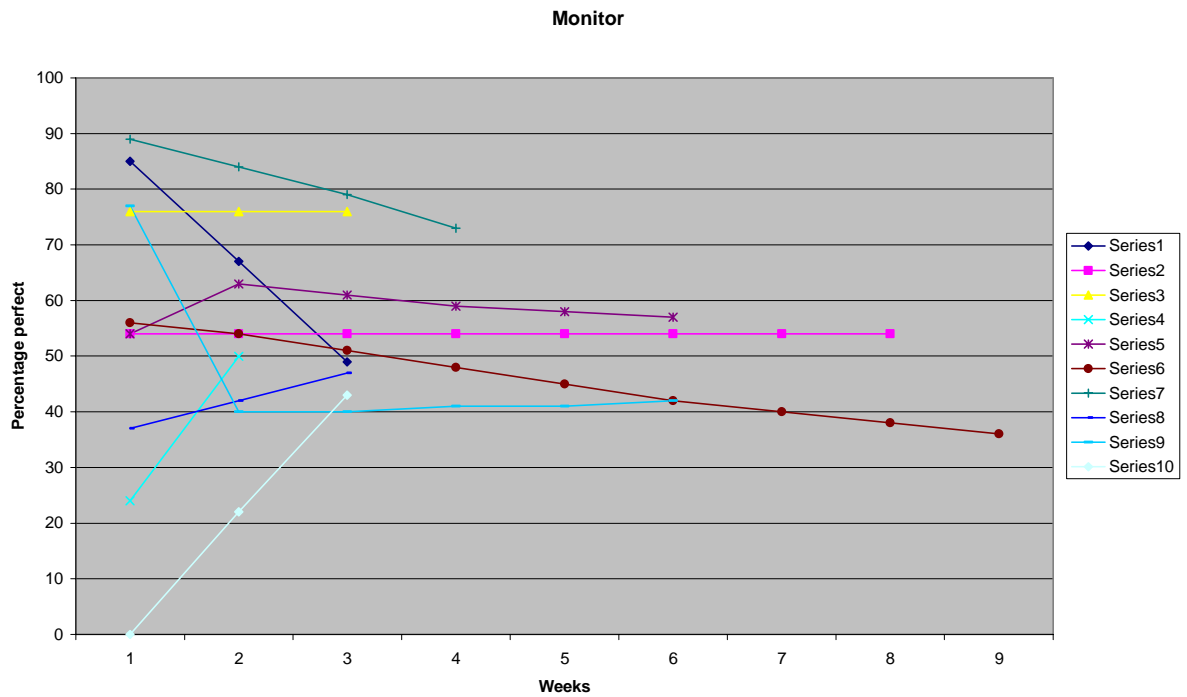
Table 4.1. Eksempel på måling af elevfremskridt.

4.1 Studieskolen



Figur 4.1. Studieskolen.

4.2 Monitor



Figur 4.2. Monitor.

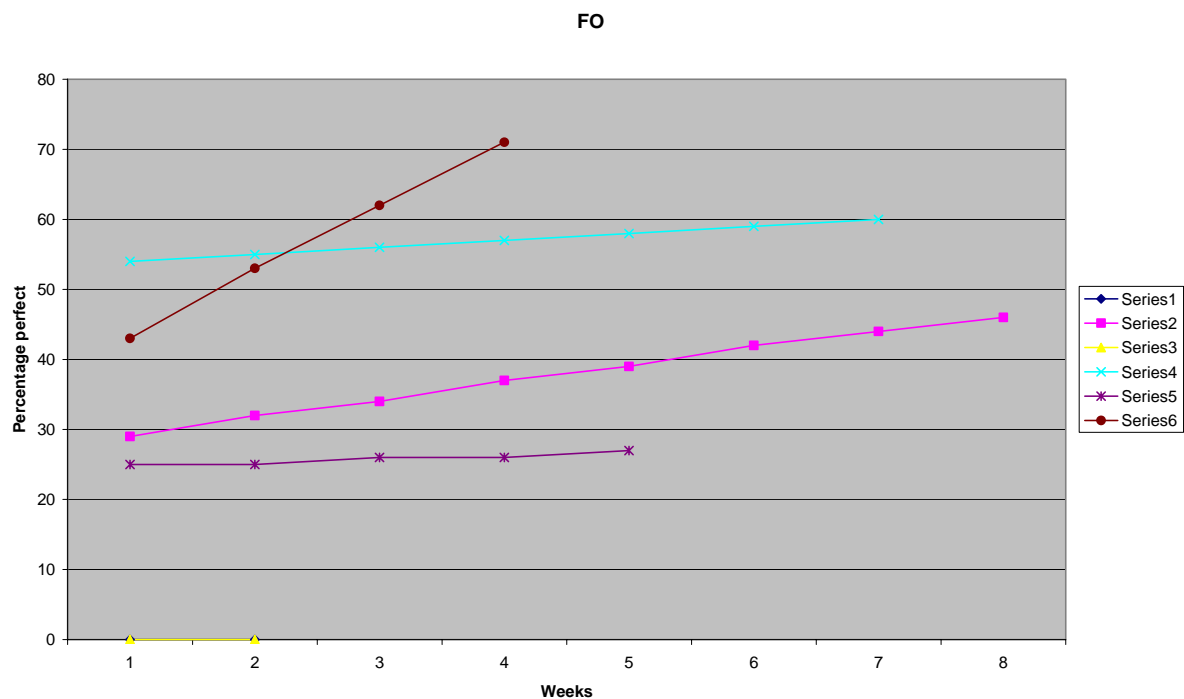
I alle graferne kan læseren identificere de elever der diskuteres under deres id som er sammensat af første bogstav i træningssteder same et løbende elevnummer. Eleven i tabel 4.1 har således id S3, dvs. elev nr. 3 fra Studieskolen. I grafen fra Studieskolen i Figur 4.1 svarer elevens nummer til "series" nummeret i grafen. Graferne bør dog printes i farver for at det bliver let at se hvilken elev hver kurve svarer til.

Graferne viser i øvrigt antal træningsuger fra starttest til slutttest (X-aksen), og elevens score i starttest, evt. mellemtest, og slutttest (Y-aksen). Det skal bemærkes, at alle andre plottede punkter end startpunkt, slutpunkt, og et evt. mellempunkt pr. elev er indsat af forfatterne for at få Excel til at vise resultaterne i liniegrafer. Disse øvrige punkter repræsenterer derfor ikke målepunkter afledt af vores data.

Grafen fra Studieskolen er i øvrigt selvforklarende.

Grafen fra Monitor er derimod ganske livlig at kigge på i begyndelsen. Det skyldes at alle anomalier blandt de analyserede elever findes hos Monitor, dvs. 4 anomalier i alt. Disse anomalier præsenteres separat i Tabel 5.1 og diskuteres i forbindelse med denne tabel.

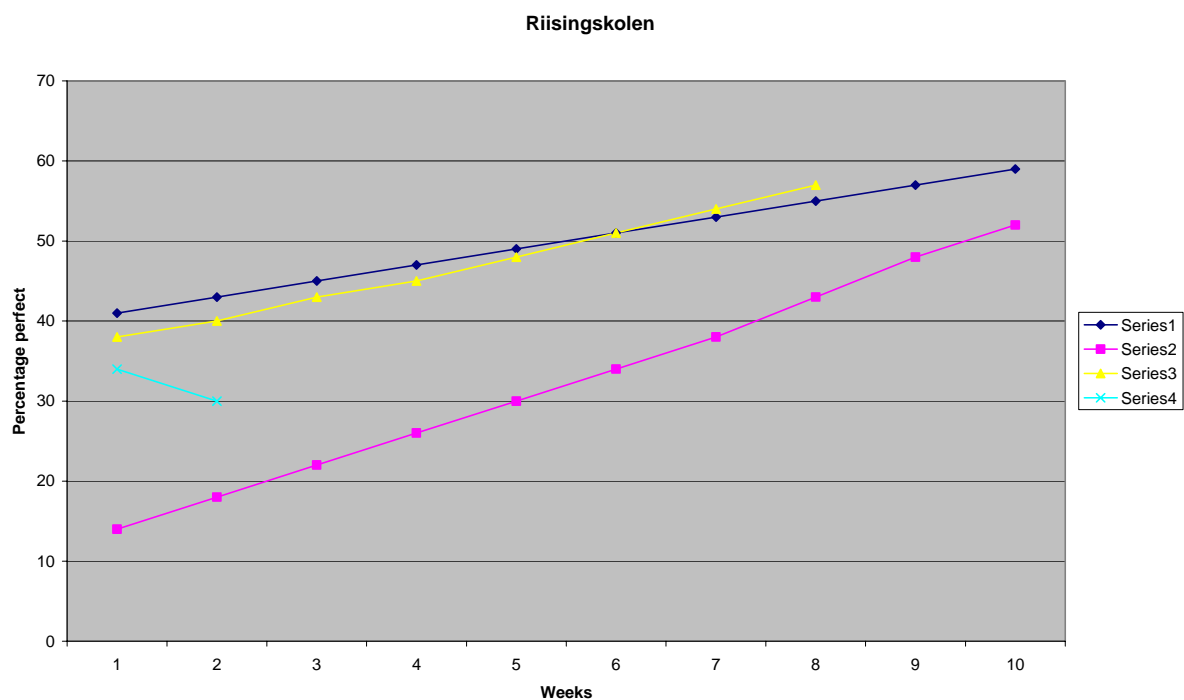
4.3 FO



Figur 4.3. FO.

Grafen fra FO er stort set selvforklarende. De to 0-0 scores man kan ane i bunden af grafen mellem uge 1 og uge 2 er fra to elever hos FO som blev udelukket pga. utilstrækkelige data men som allerede havde fået et id.

4.4 Riisingprojektet



Figur 4.4. Riisingprojektet.

Grafen fra Riisingprojektet er selvforklarende. Læg dog mærke til den korte Series4, dvs. elev R4. Sådanne elever, der kun har trænet i en uge, er marginale i vores resultater fordi de har trænet så lidt at det er vanskeligt at bygge ret meget på den trend der ses i deres fremskridt.

5 Analyse

I dette kapitel analyseres resultaterne fra de 22 elevers fremskridt med taletræneren i projektets første felttest. Det overordnede analyseprincip har været at analysere det store datamateriale i stadig dybere analyselag indtil alle de spørgsmål, vi har kunnet rejse, har fundet svar som er konsistente med alle kendte egenskaber ved datamaterialet. I fraværet af detaljer om testpersoner og træningssteder har de fundne resultater ikke kunnet korreleres med sådanne data. Vi mener på den anden side, at de nedenfor anlagte perspektiver sammenlagt er tilstrækkeligt brede og dybe til at begrunde denne rapportes hovedkonklusioner.

5.1 Overblik over resultaterne

Tabel 5.1 viser pr. elev fremskridt ifølge den primære evalueringsalgoritme (Afsnit 3.4), træningsmængden målt på forskellige måder, træningsbetingelserne ved test start og slut, og kommentarer til resultaterne. Tabellen er delt i to dele. Første del viser 18 standardtilfælde i datamaterialet hvor eleverne har gjort større eller mindre fremskridt under lignende, men ikke altid identiske, træningsbetingelser. Anden del (4 elever) viser forskellige anomalier som vi gerne vil fremhæve, samt hypoteser til at forklare dem. Tabel 5.2 forklarer forskellige udtryk i Tabel 5.1.

Fremskridt Elev	Fremskridt i % af max. mulig score	Træningsmængde	Test betingelse start/slut	Kommentarer
S1	76-70% high-drop	5 uger testindsats = 11% indsats, testord = 0 total indsats = 18%	RA, lidt V RA	Performansdrops sker også for indfødte danske talere. Beskeden træningsindsats. Kan næsten udtale enkeltord på dansk. Bemærk at sluttestbetingelsen er lidt sværere end starttestbetingelsen.
S2	33-44% low-up	5 uger testindsats = 7% indsats, testord = 0 total indsats = 13%	RA, lidt V RA	Fremskridt med meget beskeden træningsindsats, for det meste uden video.
S3	63-88% middle-high	10 uger testindsats = 9% indsats, testord = 0 total indsats = 15%	RA RA	Flot fremskridt uden brug af video. Kan udtale enkeltord på dansk.
S4	45-57% low-middle	13 uger testindsats = 33% indsats, testord = 1 total indsats = 33%	RAV RAV	Fremskridt med nogen træningsindsats.
M2	54%-54% midle-level	7 uger testindsats = 39% indsats, testord = 6 total indsats = 60%	RA RAV	Betydelig træningsindsats uden fremskridt i de analyserede testfiler. Bemærk at video ikke synes at hjælpe denne elev. Forklaringshypotese: individuelle forskelle: lærer langsomt?
M3	76-76% high-up	2 uger testindsats = 16% indsats, testord = 0	R, lidt A R	Kan udtale enkeltord på dansk. Fremskridt fordi audio ikke brugtes i sluttesten.

		total indsats = 20%		
M4	24-50% low-middle	1 uge testindsats = 6% indsats, testord = 1 total indsats = 8%	RA, lidt V R	Gode fremskridt med meget beskeden træningsindsats. Bemærk at hverken audio eller video brugtes i sluttesten.
M5	54-57% middle-up	5 uger testindsats = 33% indsats, testord = 2 total indsats = 40%	RAV RAV	Meget beskedent fremskridt med nogen træningsindsats.
M8	37-47% low-up	2 uger testindsats = 8% indsats, testord = 1 total indsats = 23%	RAV RAV	Fremskridt med beskeden træningsindsats.
M10	0-43% low-up	2 uger testindsats = 15% indsats, testord = 1 total indsats = 30%	RAV RV	Gode fremskridt med nogen træningsindsats. Bemærk at audio ikke brugtes i sluttesten.
F2	29-46% low-up	7 uger testindsats = 29% indsats, testord = 2 total indsats = 43%	RV RV	Gode fremskridt med nogen træningsindsats.
F4	54-60% middle-up	6 uger testindsats = 21% indsats, testord = 4.5 total indsats = 36%	RV RA	Gode fremskridt siden sluttesten ikke inkluderede video.
F5	25-27% low-up	4 uger testindsats = 2% indsats, testord = 0.7 total indsats = 7%	RAV RA, lidt V	Beskedent fremskridt med meget beskeden træningsindsats.
F6	43-71% low-high	3 uger testindsats = 4% indsats, testord = 0.4 total indsats = 12%	RV RV	Der var et temmelig stort "hul" i sluttestfilen. Vi kan ikke udelukke at det store fremskridt til dels skyldes at eleven undlod at udtale ord som eleven fandt vanskelige.
R1	41-59% low-middle	9 uger testindsats = 29% indsats, testord = 3 total indsats = 29%	RAV RA	Very good progress given that the final test condition did not include video.
R2	14-52% low-middle	9 uger testindsats = 38% indsats, testord = 9.5 total indsats = 38%	RA, lidt V R, lidt A	Very good progress given that the final test condition did not include video.
R3	38-57% low-middle	7 uger testindsats = 43% indsats, testord = 7.9 total indsats = 43%	R, 2/3A, 1/2V RA	Gode fremskridt siden sluttesten ikke inkluderede video.
R4	34-30% low-drop	1 uge testindsats = 11% indsats, testord = 1.5 total indsats = 24%	RAV RAV	Et drop efter en enkelt uges træning kan være uden betydning.

Anomalier Elev	Beskrivelse	Træningsmængde	Test betingelse start/slut	Kommentarer
M1	85-49% high-low drop	2 uger testindsats = 6% indsats, testord = 2 total indsats = 13%	R R	Meget beskeden træningsindsats. Det store drop er uforklaret. Bemærk at eleven kun bruger tekst. Startscoren på 85% antyder at eleven kan udtale enkeltord på dansk. Forklaringshypoteser: useriøs slutttest? Forskellige personer i de to tests?
M6	56-36% middle-low drop	8 uger testindsats = 20% indsats, testord = 2.4 total indsats = 28%	RAV RA	Synes at være klart tilfælde af hvad der sket når eleven går fra at have videostøtte til at undvære denne. Forklaringshypotese: video ikke brugt i slutttest.
M7	89-73% high-drop	3 uger testindsats = 11% indsats, testord = 0 total indsats = 15%	R, lidt V R, lidt V	Performansdrops sker også for indfødte danskere. Beskeden træningsindsats. Kan udtale enkeltord på dansk. Performansdroppet er iøvrigt uforklaret.
M9	77-42% high-low drop	5 uger testindsats = 30% indsats, testord = 4 total indsats = 40%	RV R, lidt A	Synes at være klart tilfælde af hvad der sket når eleven går fra at have videostøtte til at undvære denne. Forklaringshypotese: video ikke brugt i slutttest.

Tabel 5.1. Testresultater, træningsmængde og træningsbetingelser pr. elev.

5.1.1 Forklaringer til Tabel 5.1

Term	Forklaring
A	Eleven brugte audio for at høre ordene i testsekvensen før de udtaltes
Elev id	Elever fra Studieskolen, Monitor, FO, og Riisingprojektet er mærket henholdsvis S(n), M(n), F(n), og R(n)
Fuldt trænings- og testpensum	Antagen træningsmængde = 450 ord pr. uge i 10 uger = 4500 ord
Indsats, testord	Antal gange eleven har trænet og testet med testordene mellem start- og slutttest
R	Eleven brugte skreven tekst for at læse ordene i testsekvensen før de udtaltes
Testindsats	Estimeret antal ord udtalt fra starttest til slutttest i % af fuldt træningspensum
Total indsats	Estimeret antal ord udtalt under al træning og testning i % af fuldt træningspensum. Eleverne plejer at gå igennem alle 450 testord gentagne gange men der er individuelle forskelle i hvor systematisk de gør det og hvilken systematik de følger
Uger	Antal uger mellem første og sidste testdato
V	Eleven brugte video for at se og høre ordene i testsekvensen før de udtaltes

Tabel 5.2. Forklaringer til Tabel 5.1.

5.1.2 Kommentarer til Tabel 5.1

Tabel 5.1 viser i grunden to centrale sæt af informationer. For det første opsummerer den pr. elev elevens kvantitative fremskridt i de målte tests, dvs. de som også vises i graferne i Afsnit 4. Af Tabel 5.1 kan man også se under hvilke testbetingelser hver test er udført (Søjle 4). Højre søjle i Tabel 5.1 kommenterer om elevernes fremskridt. For det andet viser tabellen en

række centrale målinger af elevernes træningsindsats (Søjle 3). Hvert af disse mål forklares i Tabel 5.2.

Derudover er Tabel 5.1 opdelt i en (1) hovedgruppe af elever som gennemfører deres tests under nogenlunde samme testbetingelser, bortset fra at godt halvdelen af dem i deres endelige test valgte lidt vanskeligere testbetingelser; og (2) en gruppe på 4 anomalier i materialet.

5.1.3 Anomalierne i Tabel 5.1

Anomalierne er følgende:

Elev **M1** er en ægte anomali. Vi har ingen forklaring på det målte store drop i præstation fra 85% til 49% i testen. Kontrolmålingerne viser præstationer imellem disse ekstremer. Men en vigtig pointe er følgende. Hvis en elev er i stand til at præstere 89% korrekt med taletræneren blot ved at læse de ord som skal udtales, som tilfældet er for M1, så taler eleven korrekt dansk og har ikke brug for at tale med taletræneren. Eleven er på, eller tæt ved, sin maksimumpræstation allerede, og enhver måling af elevens frem- eller tilbageskridt svarer til at måle daglige udsving i udtale hos indfødte danskere.

Mere generelt forekommer det os, at alle elever som præsterer over 70% under en test *alene vha. at læse testordene*, allerede må være tæt på deres maksimum. Derfor er fremskridtsmålinger på disse elever ret omsonst. Det er faktisk urimeligt at anvende disse elevers frem- eller tilbageskridt i målinger af gennemsnitlige fremskridt hos alle elever. Det bemærkes i denne forbindelse at hovedgruppen i Tabel 5.1 også rummer to elever med denne profil. Dog bruger S1 også audio som testbetingelse, hvorfor denne elev alligevel hører til i hovedgruppen. Den anden elev, M3, er også i hovedgruppen fordi denne elev ikke kun bruger læst tekst i starttesten, men som man kan se fra tabellen er denne elevs tilhørsforhold til hovedgruppen marginal.

Af ovenstående grunde er også **M7** en anomali. De 89% i starttesten under brug af en smule video viser, at M7 er tæt på sit maksimum og må antages at kunne udtale dansk stort set korrekt. M7s tilbagefald under testforløbet svarer derfor til hvad indfødte danskere kunne præstere.

I nogle af beregningerne nedenfor vil vi se bort fra M1 og M7 fordi disse to elever et tæt på deres maksimum allerede i starttesterne.

De to øvrige anomalier er **M6** og **M9**. Deres kraftige drop i præstation forklarer vi med at de begge i deres slutttest undlader at bruge video medens de brugte video i starttesten. Disse to eksempler viser den endog meget kraftige effekt af at ændre testbetingelserne fra brug af video til at klare sig uden video. Det ville være helt misvisende at tage M6 og M9 med i hovedgruppen, fordi hovedgruppen viser fremskridt under relativt uændrede, men ikke helt uændrede, testbetingelser.

5.2 Træningsmængde og resultater

Tabel 5.3 opgør elevernes gennemsnitlige træningsmængde på forskellig vis.

Det fremgår at eleverne i gennemsnit har trænet i 5.3 uger mellem deres første og sidste test. Deres samlede træning målt i antal udtalte ord er 26.7% af et fuldt træningspensum på 4500 ord. Vi kan med andre ord klart se at elevernes som helhed har trænet langt mindre end et fuldt træningspensum.

Mellem eleverne er der meget store forskelle i den samlede træningsmængde, nemlig, som det fremgår af Tabel 5.1, fra 7% til 60% af et fuldt pensum.

Disse observationer betyder, at *denne rapport's resultater må betragtes som helt foreløbige når det drejer sig om at evaluere taletrænerens fulde potentiale til selvtræning*. Vi har stadig intet datamateriale som for et tilstrækkeligt antal elever viser effekten af at gennemføre et

fuldt træningspensum. Vi har faktisk ikke data fra et fuldt træningspensum fra en eneste elev. Det rejser spørgsmålet om det er muligt fra de foreliggende resultater at ekstrapolere sig til, eller på anden vis estimere, indlæringseffekten af at gennemføre et fuldt træningspensum.

	Antal elever	Træningsuger mellem tests	Træning mellem tests i % af fuldt pensum	Gentagelse af testsekvens i testperioden	Træning totalt i % af fuldt pensum
Hovedgruppen	18	98, snit 5.44	354%, snit 19.67%	41.5, snit 2.31	492%, snit 27.33%
Anomali-gruppen	4	18, snit 4.50	67% snit 16.75%	8.4, snit 2.10	96%, snit 24%
Alle	22	116, snit 5.27	421, snit 19.13%	49.9, snit 2.26	588%, snit 26.73%

Tabel 5.3. Træningsmængder, oversigt.

Tabel 5.3 viser at eleverne i gennemsnit har trænet 19.1% af et fuldt træningspensum *mellem* deres første og sidste test. Hovedgruppen på 18 elever i Tabel 5.1 har haft et gennemsnitligt fremskridt på 14.2% mellem starttest og sluttet. Vi foreslår en lille korrektion til dette gennemsnit ved at udelukke de to elever hvis resultater ligger højt i begge tests, dvs. over 70%. Disse elever er formentlig allerede omkring deres maksimale præstation ved starten af testforløbet. Med denne korrektion får vi et gennemsnitligt fremskridt for de resterende 16 elever i hovedgruppen på 16.4%.

5.2.1 Træningsprisen for 10% fremgang

Baseret på ovenstående tal kan vi regne ud, at prisen for **10% fremgang i testscore** i absolutte procenttal adderet til deres startscore for de 16 elever er en træningsindsats på knap **12% af et fuldt pensum**. Dette er et ganske godt resultat, må man sige, et som ifølge Tabel 5.3 i gennemsnit kan gennemføres på knap 3.5 uger med et ganske moderat træningspensum per uge. Ydermere bemærkes det at dette resultat ifølge de indsamlede feltdata er et **minimumsresultat**. 12% af fuldt pensum giver faktisk mere fremgang end 10% når man også tager højde for testbetingelserne. Mere end halvdelen af hovedgruppen af testpersoner gjorde sluttet sværere for sig selv ved at vælge vanskeligere testbetingelser end i starttesten (Tabel 5.1). Desværre er det ikke muligt i øjeblikket at kvantificere denne yderligere fremgang i udtalekompetence. Det kræver flere data fra elever som har testet sig selv med samme testsekvens i forskellige træningsbetingelser.

Desuden er det meget vigtigt at forstå ovenstående gode resultat i lyset af testbetingelserne. Så, for at være mere præcis, er resultatet at eleverne *i samme eller lidt sværere testbetingelser* opnår deres *første 10% fremgang* vha. *12% af det fulde kursuspensum*.

5.2.2 Lineært fremskridt?

Hvor langt kan man ekstrapolere fra det gode resultat vist i Afsnit 5.2.1, givet at vi ikke har data fra fuldt gennemførte træningskurser hvor 4500 ord er trænet og testet? Det kan vi ikke sige for øjeblikket, fordi vi ikke har fuldt gennemførte træningspensum for et passende antal elever. Faktisk har ingen elev gennemført det fulde træningspensum endnu.

To ekstremer er i denne forbindelse ret klart forkerte. Det første ekstrem er at fremskridt med taletræneren er lineært. Dette er naturligvis forkert, ikke blot fordi man ikke kan nå et resultat højere end en 100% perfekt score, men også fordi erfaringen viser, at voksne som lærer dansk som andetsprog ofte stopper i deres fremgang på et sted hvor man stadig kan høre en mere eller mindre tydelig accent. Vores baselining af taletræneren med danskere viste, at indfødte

danskere opnåede en testscore i den vanskeligste testbetingelse (læse) på mellem knap 80% og tæt på 95%.

Et skøn kunne derfor være, at det gælder om at få elever med dansk som andetsprog op på min. 70%. Derpå må de gerne nå et plateau for alle vil nu kunne forstå deres danske udtale selv om der er en accent.

Det andet ekstrem er at fremskridt med taletræneren hurtigt vil nå et plateau efter de første 10-20% fremgang. Det mener vi med resultaterne i denne rapport at have demonstreret er forkert. Der er allerede en del elever som med et relativt beskedent træningspensum viser en fremgang på 30% eller lignende.

Sandheden ligger derfor nok et sted mellem de to ekstremer. Vi vil også faktisk være i stand til at kunne finde sandheden, i det mindste som en rimelig approximation, når vi har et passende antal elever som har gennemført fulde træningspensum.

Igen skal det understreges, at diskussionen ovenfor angår elever som tester under samme eller lidt sværere testbetingelser. Det betyder f.eks., at en elev som med 70% fuldt træningspensum når fra 32% til 72% under den letteste testbetingelse, dvs. (læse + høre + se og høre) før udtale af hvert ord, stadig kan have en del at lære før eleven når sit topplateau i dansk udtale. Derfor ville det være særdeles nyttigt at have en omregningsfaktor, baseret på tilstrækkelige data, som viser hvor godt en elev som præsterer en X% score i (læse + høre + se og høre) – testbetingelsen, i gennemsnit vil score hvis testbetingelsen ændres til en af de sværere (læse + høre) eller (læse). Vi mangler data til at beregne en approximation til en sådan faktor eller til sådanne faktorer.

5.2.3 Konklusion om fremskridt som funktion af træningsmængde

Efter analyse af de foretagne feltafprøvninger af taletræneren har vi lært nogle centrale og grundlæggende kendsgerninger, nemlig at taletræneren giver eleverne gode og hurtige fremskridt på 15-20% i gennemsnit, i absolutte procenttal som adderes til deres udgangsscore, givet at eleverne tester under de samme eller lidt sværere testbetingelser.

Vi har også konstateret at der er to meget vigtige faktorer som vi ved alt for lidt om pt. Den første er en fremskridtsestimatfunktion som tillader os at estimere gennemsnitsfremskridtet for en elev med et fuldt træningspensum, givet at eleven har gennemført X% af det fulde pensum. Den anden er en omregningsfunktion som tillader os at estimere gennemsnitsfremskridtet for en elev som har gennemført X% af det fulde pensum og scoret Y% fremgang i samme testbetingelse, og hvor vi ønsker at beregne hvor stort fremskridt eleven vil kunne gøre med et fuldt træningspensum i en vanskeligere testbetingelse.

5.3 Individuelle forskelle og forskelle i modersmål

Vi har i skrivende stund ikke adgang til personlige data om eleverne. Det er klart at disse er yderst forskellige mht. modersmål, længden af deres hidtidige ophold i Danmark, omfanget af deres hidtidige sprogskoleuddannelse, evner til hurtigt at tilegne sig udtale af et nyt sprog, alder, osv. Derfor kan vi ikke forvente i det forhåndenværende materiale, når det ikke grupperes efter parametre som de netop nævnte, at finde nogen særlig korrelation mellem træningsmængde mellem starttest og sluttest, og fremskridt i dansk udtale. Vi kan dog notere fra Tabel 5.1 at hurtige fremskridt ikke er begrænset til elever som starter med at score lavt, hvilket forekommer ganske lovende. Tabellerne 5.4 og 5.5 viser det netop sagte.

R3	19%	43%
M2	0%	39%
R2	38%	38%
S4	12%	33%

M5	3%	33%
F2	17%	29%
R1	18%	29%
F4	6%	21%
M10	43%	15%
R4	-4%	11%
S3	25%	9%
M8	10%	8%
S2	11%	7%
M4	26%	6%
F6	28%	4%
F5	2%	2%
Elev	Fremskridt	Test indsats

Tabel 5.4. Fremskridt i relation til træningsmængde.

M10	43%	15%
R2	38%	38%
F6	28%	4%
M4	26%	6%
S3	25%	9%
R3	19%	43%
R1	18%	29%
F2	17%	29%
S4	12%	33%
S2	11%	7%
M8	10%	8%
F4	6%	21%
M5	3%	33%
F5	2%	2%
M2	0%	39%
R4	-4%	11%
Elev	Fremskridt	Test indsats

Tabel 5.5. Træningsmængde i relation til fremskridt.

Det er klart fra Tabel 5.1 at der er særdeles store individuelle forskelle i elevernes fremskridt. Som det ene ekstrem kan nævnes elev M2 som ikke gør fremskridt overhovedet til trods for at netop denne elev (1) træner mere end alle andre elever i datamaterialet, nemlig 60% af fuldt pensum; (2) træner 39% af fuldt pensum mellem start- og sluttest; og (3) endog gennemfører sluttesten i en normalt lettere betingelse (læse + høre + se og høre) end starttesten (læse + høre). Datamaterialet viser kun en enkelt anden elev som gennemfører sluttesten i en lettere testbetingelse end starttesten. Alle andre elever gennemfører sluttesten i samme eller i en vanskeligere testbetingelse end starttesten. M2 ligger på en stabil 54% testscore gennem hele det målte træningsforløb.

Som det andet ekstrem kan f.eks. nævnes S3 som med kun 9% af fuldt træningspensum forbedrer sig fra 63% til fantastiske 88% og under samme træningsbetingelse (læse og høre). Da forfatterne til denne rapport endnu ikke har fået adgang til detaljer om hver enkelt elev kan vi ikke komme ind på hvor stor en del af disse enorme forskelle der har at gøre med elevens

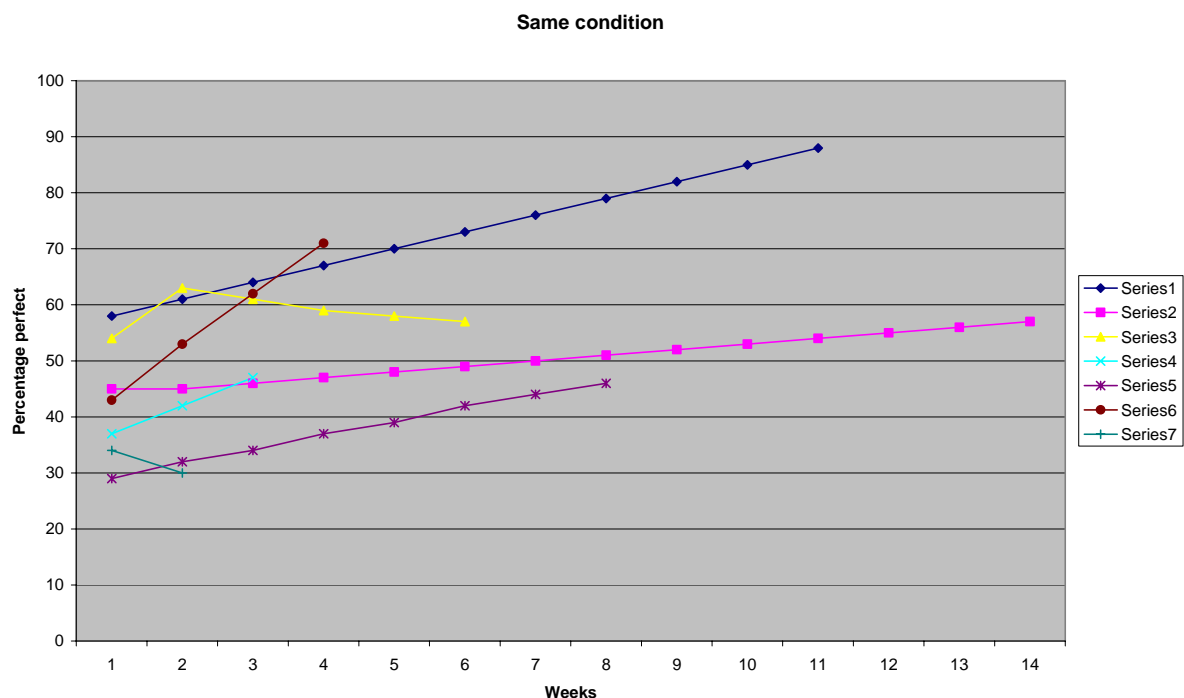
modersmål. Vi ved at S3's modersmål er polsk men vi kender ikke M2's modersmål. Derfor kan vi endnu ikke sige noget om betydningen af elevernes modersmål i relation til deres fremskridt eller stagnation i dansk udtale.

Men det er klart at individuelle forskelle også er meget vigtige. Nogle elever har et godt sprogmemne medens andre er mere sprogtungnemme. Så længe vi ikke har modersmålsoplysninger om eleverne giver det dog ikke megen mening at forsøge at separere disse to vigtige faktorer. Vi behøver nok også data fra flere elever end vi har pt. for at kunne gøre dette med nogen antydning af sikkerhed i konklusionerne.

For øjeblikket er det imidlertid nærliggende at anbefale, at taletrænerelever monitoreres for at se om de i løbet af f.eks. 5 uger og med gennemførelsen af 30-50% af pensum har gjort væsentlige fremskridt eller ikke, igen under hensyntagen til testbetingelserne. Har de ikke gjort væsentlige fremskridt under samme testbetingelser, så bør man nok sætte en sproglærer på dem for evt. gennem individuel rådgivning at kunne "afblokere" manglende fleksibilitet hos dem i deres træning af dansk udtale. Der kan også være elever som bare ikke kan gøre fremskridt, men det har vi absolut ingen evidens for at antage i lyset af vores feltdata.

5.4 Elever med samme testbetingelser

Figur 5.1 viser fremskridtene hos de 7 elever i normalgruppen i Tabel 5.1 som valgte de samme testbetingelser i start- og sluttesten.



Figur 5.1. Elever med samme testbetingelser i start- og sluttest.

Det interessante er at disse elevers gennemsnitlige fremgang er 14%, dvs. lidt lavere end gennemsnittet for alle eleverne i normalgruppen. Umiddelbart kunne dette se ud til at vise, lidt selvmodsigende udtrykt, at en test *ikke* bliver sværere ved at vælge en sværere testbetingelse. Vores hypotese er imidlertid, at stort set alle elever er sig pinligt bevidste om at der er forskel i sværhed mellem testbetingelserne. Eleverne i Figur 5.1 består derfor for en dels vedkommende af elever som ikke har mod på at teste sig selv i en sværere testbetingelse, eller som har sat sig det mål at komme så højt op som muligt i samme testbetingelse. De elever som tester sig selv under sværere betingelser til slut er derimod de som har sat sig et højere mål med deres træning.

5.5 Kontrolmålinger

Vi fokuserer i denne rapport på at sammenligne fremskridt pr. elev i at udtale *samme ordsekvens* under lignende træningsbetingelser, jv. Afsnit 3.4. Kravet om at finde samme ordsekvenser til sammenligning i en elevs logfiler betyder, at (1) de sammenlignede testresultater ofte ikke stammer fra meget tidligt i træningen resp. meget sent i træningen. Samtidig udgør de logfiler vi har målt på kun en mindre del af alle en elevs logfiler. Det er derfor i høj grad muligt, at (2) de to (eller tre) testmålepunkter pr. elev kunne give et misvisende billede af elevens fremskridt under hele træningsperioden, specielt givet de mange og for det meste for os ukendte uafhængige variable som kunne påvirke hver enkelt test, jf. Afsnit 3.8.

Af disse to grunde har vi foretaget 3 kontrolmålinger pr. elev på nye logfiler som vi ikke har brugt før, jf. Afsnit 3.6. Kontrolmålingerne er foretaget på den tidligste logfil som ikke er målt før, den seneste logfil som ikke er målt før, og en tilfældigt valgt logfil mere eller mindre midt imellem de to (i tid). Disse 3 nye logfiler vil normalt ikke vise elevens præstation på samme ordsekvenser som brugt med den primære evalueringsalgoritme (Afsnit 3.4). Men de vil vise en tidslig progression for eleven som efterfølgende kan sammenlignes med den allerede målte progression i henhold til den primære evalueringsalgoritme.

De primære og de sekundære (samplede) testmålinger pr. elev indsattes i en tabel som også viste de testbetingelser, eleven havde haft under de i alt 5 testmålinger pr. elev repræsenteret i tabellen. Endelig kommenteres i tabellen pr. elev på forholdet mellem de to serier af resultater. Et eksempel vises i Tabel 5.3.

Elev	Testscores start-slut %	Test betingelser	Samplede scores %	Test betingelser	Sammenligning samplede scores/ testscores Andre kommentarer
S3.	63 88% middle-high	RA RA	60 58 80	RA RA RA	Bekræfter test

Tabel 5.3. Eksempel på sammenligning mellem primære testscores og samplede testscores.

Vores hovedhypoteser vedrørende hvad sammenligningen mellem de to testserier, hovedtestserien og sample testserien, ville vise, var:

- større spredning i sample testserien fordi denne generelt dækker et længere testforløb;
- i de fleste tilfælde, samme fremskridtsrate i hovedtest og sample test;
- højere slutscores i sample testen end i hovedtesten i en række tilfælde, givet samme testbetingelser, fordi sample testen ofte fandt sted senere i træningsforløbet;
- de samme effekter af forskellige testbetingelser som vi fandt i hovedtesten.

Hvad vi fandt var følgende:

- klart større spredning i sample testen;
- samme fremskridtsrate i hovedtest og sample test i 16 tilfælde;
- klart bedre præstation i sample test end i hovedtest i 3 tilfælde;
- klart dårligere præstation i sample test end i hovedtest i et enkelt tilfælde;
- bekræftelse på scoringsmønstret i de anomali-tilfælde i Tabel 5.1 som illustrerer effekten af test med og uden video: 2 tilfælde.

Som konklusion viser sample testen (1) *samme resultatmønstre* som hovedtesten samt (2) at en række elever i det samlede træningsforløb opnåede *større udtaleforbedringer* end

hovedtestens tal viser. Mere overordnet forekommer det at vi kan konkludere, at langt de fleste elever har trænet målbevidst og seriøst med taletræneren. Havde de ikke det, ville sammenligningen mellem resultaterne af hovedtesten og sample testen havde vist langt flere inkonsistenser og tilsyneladende tilfældige fluktuationer end tilfældet er. Dette er i sig selv et betydningsfuldt resultat som viser, at taletræneren i sin nuværende udformning faktisk inspirerer næsten alle elever til at gøre deres bedste for at opnå et godt resultat.

5.6 Træningsbetingelser - igen

Det har for os været en opdagelse i datamaterialet at se i hvor høj grad elevernes score afhænger af testbetingelserne. Eleverne selv synes at have en klar forståelse af under hvilke testbetingelser det bliver sværere for dem at opnå deres bedste score. Med undtagelse af blot to tilfælde i sammenlagt $5 \times 22 = 110$ målepunkter gælder det for alle elever, at de *enten* tester sig selv ved start og slut under *de samme* testbetingelser eller de tester sig selv til slut under *sværere* testbetingelser. De elever der gør det sidstnævnte demonstrerer alle fremskridt alligevel, med undtagelse af de to anomalier som vælger drastisk at droppe video fuldstændigt i sluttesten. Desuden kan vi se at de fleste elever som præsterer en høj startscore ikke bruger video eller kun bruger video i begrænset omfang. Indfødte danskere har heller ikke brug for video, eller for audio, for den sags skyld. For dem er det nok at læse det ord der skal udtales og så udtale det i henhold til deres mentale model af det skrevne ords udtale.

Vi drager to konsekvenser:

- en bekræftelse af vores hypotese om rangordenen i sværhed af de forskellige testbetingelser;
- langt de fleste elever har faktisk trænet seriøst og målbevidst.

5.7 Anbefalinger

5.7.1 Brugermanual og elevinstruktion på træningsstedet

Eleverne bør i brugermanualen opfordres til at teste sig selv med samme ordsekvens under forskellige testbetingelser. Når de begynder at kunne klare sig uden video bør de gøre det, idet de derved kommer op på et højere niveau i udtaletræner-”computerspillet”. Deres træningsmål er ikke alene at komme op i ”high performance” –området over 70%, men i sidste ende at komme derop uden brug af video forud for udtale af ordene i testsekvensen. På grundlag af de data som er analyseret i denne rapport synes det i sammenligning mindre vigtigt om eleverne i de svære tests bruger både tekst og audio, tekst alene, eller audio alene. Men eleverne bør opfordres til selv at udforske disse forskelle.

Eleverne bør så vidt muligt træne det fulde anbefalede pensum, dvs. 4500 ord svarende til at alle ord i taletræneren er udtalt 10 gange.

Eleverne bør forsøge at træne det anbefalede pensum over en periode på ca. 10 uger. Dette skulle være praktisk muligt som led i deres dansktræning. Som det er fremgået ovenfor har de analyserede elever brugt forholdsvis længere tid på deres træning end hvad der svarer til at gennemføre det fulde pensum på 10 uger.

Hvad angår personlig elevinstruktion på træningsstederne så vi i Afsnit 2.1 store forskelle mellem træningssteder, hvor mange elever havde ”snuset” til taletræneren men kun få havde brugt den intensivt nok til at levere data til vores dataanalyse, og træningssteder hvor hele 57-67% af eleverne havde leveret brugbare data til analysen. En nærliggende forklaring er, at man på de sidstnævnte træningssteder har taget sig godt af elevernes træning og har motiveret dem til at blive ved med at bruge taletræneren.

Det er klart at vi endnu befinder os på et meget tidligt stadium i indhøstning og bearbejdning af erfaringer med taletrænerens brug i praksis i felten. Men det forekommer umiddelbart indlysende, og som sagt også sandsynligt fra de data vi har set allerede, at taletræneren i fremtiden bør integreres fuldt i træningsstedernes undervisning i dansk, hvilket sandsynligvis vil kunne medføre en mere ensartet og effektiv motivation af eleverne til at bruge taletræneren intensivt i deres selvtræning.

5.7.2 Taletrænersystemet

Arbejdet er i gang på at modificere taletræneren så eleverne selv kan informere systemet om de vil træne eller teste. Dette vil lette analysen af deres resultater, samtidig med at eleverne selv straks vil få endnu bedre feedback på deres fremskridt end de gør med den nuværende version af systemet. Vi overvejer også i denne forbindelse at give eleverne valget mellem flere forskellige testbetingelser når de vælger at teste sig selv. Dette vil igen understrege betydningen af at teste sig selv under forskellige betingelser, og det vil forhåbentlig give os bedre data om de kvantitative fremskridtseffekter af at samme elev tester sig selv under klart forskellige betingelser. Som vi har set er der mange elever som tester sig selv under blandede betingelser, såsom (læse *og* høre *og* en smule video). Disse blandede testbetingelser vanskeliggør den efterfølgende analyse af elevernes fremskridt.

6 Konklusioner

Den første felttest med taletræneren har givet et stort og særdeles værdifuldt materiale om taletrænerens brug i praksis på træningsstederne samt om taletrænerens reelle nytte til at forbedre dansk udtale af enkeltord hos personer med et andet modersmål end dansk.

Vi kan konstatere, at stort set alle elever på træningsstederne har gennemført seriøs og målbevidst træning og at de har været klart bevidste om virkningen på deres testpræstationer af at teste under forskellige testbetingelser. Dette viser, at taletræneren faktisk virker i praksis, og gør det sandsynligt at taletræneren understøtter elevernes motivation til at træne med systemet.

Vi kan også konstatere, at eleverne som helhed har trænet for lidt til at give os det ideelt ønskede udbytte af evalueringen af taletrænerens brug og nytte på træningsstederne. Eleverne har i gennemsnit gennemført godt 1/4 af det fulde træningspensum på 4500 ord. Da stort set alle træningssteder var i gang pr. 1.2.2005 repræsenterer vores datamateriale mindst ca. 12 ugers data fra de leveringsdygtige træningssteder. Mange elever stoppede dog deres træning væsentligt tidligere end efter 12 uger. Da vi ikke har direkte adgang til eleverne på træningsstederne må vi fremover finde på andre midler til at forsøge at intensivere elevernes træning. Det skal dog i denne forbindelse siges, at feltræningen med taletræneren fortsætter frem til sommeren 2005, så vi håber til den tid at have fået det ønskede materiale om (næsten) fuldt gennemførte træningsperioder for i det mindste nogle elever.

Det er nu klart at eleverne gør gode fremskridt med taletræneren. I gennemsnit har de 16 elever i hovedgruppen som ikke allerede taler et godt dansk, opnået et fremskridt, som minimum, på 16.4% i absolutte procenttal over deres startscore, og med en gennemsnitlig træningsindsats på ca. 20% af det fulde anbefalede træningspensum. Dette gennemsnit dækker over meget store individuelle forskelle i elevernes fremskridt, hvilket igen betyder at mange elever har gjort væsentligt større fremskridt end gennemsnittet viser. Vi vil anse det for et konservativt skøn, at elevernes gennemsnitlige fremgang ved brug af taletræneren med en væsentlig større træningsindsats vil kunne nå mindst 40% absolutte procent over deres starttestscore, således at en elev som i starttesten opnår en score på 30% i slutttesten vil opnå en score på 70% under de samme testbetingelser som i starttesten.

Vi mener, at de her fremlagte resultater giver et ret overbevisende grundlag for at antage, at taletræneren er et nyttigt redskab til selvtræning i dansk udtale for personer med et andet modersmål end dansk. De forbehold, vi er nødt til at tage, er at vi gerne vil se resultaterne fra et passende antal elever som har gennemført det fulde træningspensum, og at vi gerne vil se på en række sammenhænge mellem elevernes fremskridt under træningen og deres personlige data, herunder også brug af lyttetræneren.

Med disse forbehold mener vi at denne rapport etablerer et forsvarligt grundlag for at fortsætte udtaletrænerprojektet med henblik på dansk taletræning på sætningsniveau.