

Tonem 1 eller 2 eller 1,5?

Arild Noven (arild.noven@gri.no), Per Arne Larsen (pelarsen@gri.no),
Bente Moxness (bente.moxness@gri.no), Kolbjørn Slethei
(kolbjorn.slethei@gri.no)
Voss International Language Technology A/S

Samandrag

Artikkelen presenterer eit eksperiment utført ved Nordisk Språkteknologi A/S der ein vurderte nytten av eit nøytralt tonem som erstatning for tonem 1 eller 2 i ein norsk talesyntetisator. Arbeidshypotesen var at val av dialekt er heilt avgjerande for utfallet, og Ålesunds-dialekten vart vald avdi han har ein minimal skilnad mellom tonem 1 og 2. Eit nøytralt tonem 1,5 vart konstruert som ein interpolasjon mellom tonem 1 og 2, og 2 testsetningar, kvar med ulike kombinasjonar av tonem 1, 1,5 og 2, vart presentert for ei gruppe utvalde forsøkspersonar. Forsøkspersonane vart så bedne om å vurdere kvaliteten på testsetningane langs ein 5 punktsskala. Resultata viste at tonem 1,5 vart akseptert som fullgod erstatning for dei ekte tonema.

1. Innleiing

Det å konstruere ein talesyntetisator byr på mange ulike typar utfordringar frå dei reint tekniske til dei meir lingvistiske. Ein stor del av det prosesseringsarbeidet som vert utført frå skriven tekst til ferdig lyd, er av lingvistisk og fonetisk art. Dei skandinaviske språka har mange lingvistiske og fonetiske særtrekk som gjer eit slikt prosjekt vanskelegare enn for mange andre europeiske språk. Tonelagsskilnad, utstrekt bruk av komposita og den komplekse morfologien i slike ord er mellom dei tinga som gjer dette arbeidet krevjande.

I 1984 kom den første versjonen av ein norsk talesyntetisator – den såkalla INFOVOX-syntetisatoren - som resultat av eit samarbeidsprosjekt mellom KTH i Stockholm og Universitetet i Bergen. Under utviklinga av denne syntetisatoren, vart det klart at særleg to problemområde var avgjerande for kvaliteten, nemleg trykkplassering og val av tonem. Sidan dei aller fleste norske dialektar er tonemiske, og bruk av korrekt tonem er ein viktig faktor i ein talesyntetisator av høg kvalitet, var det avgjerande å finna reglar som styrde tildeling av rett tonem til dei aktuelle orda.¹ Imidlertid viste dette seg å vera eit ikkje-trivielt problem sidan både leksikalske og bøyingsmorfologiske faktorar spelar inn. (Sjå også Carlson *et al.* (1982) og Carlson & Granström (1986).) Tabell 1 viser litt av denne kompleksiteten ² (Tonem 1 og 2 er markert med ' og ")

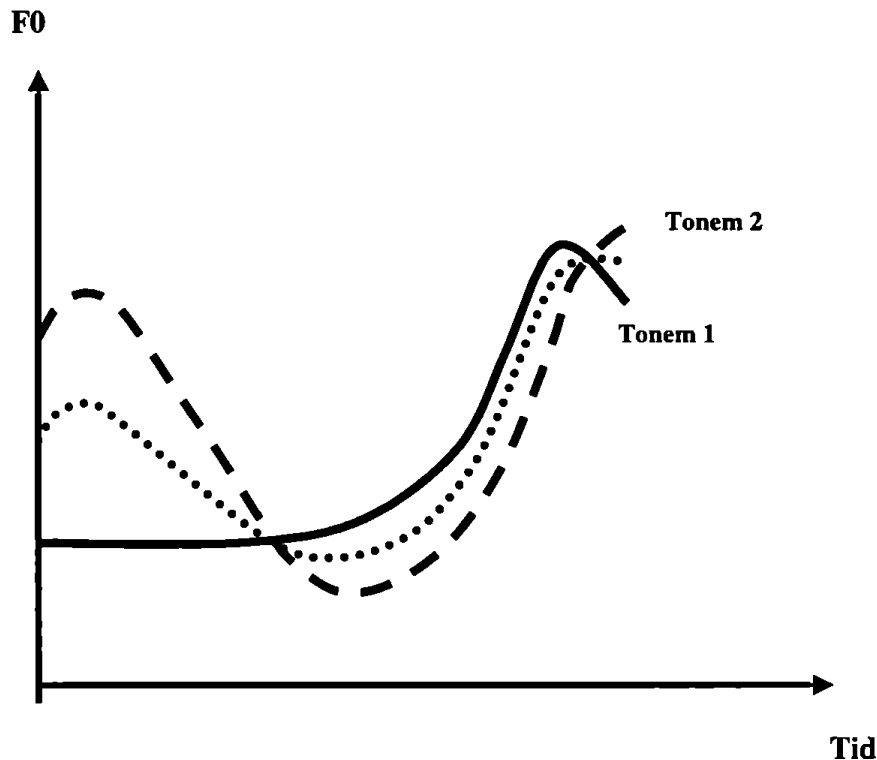
Bøyingsform	Struktur	Transkripsjon	Struktur	Transkripsjon
Infinitiv	løp-e	/'lø:pe/	hopp-e	/'hɔpe/
Presens	løp-er	/'lø:per/	hopp-er	/'hɔper/
Preteritum	løp-	/'lø:p/	hopp-et	/'hɔpet/
Ubest. sing.	løp-	/'lø:p/	hopp-	/'hɔp/
Best. sing.	løp-et	/'lø:pe/	hopp-et	/'hɔpe/
Ubest. plur.	løp-	/'lø:p/	hopp-	/'hɔp/
Best. plur.	løp-ene	/'lø:pene/	hopp-ene	/'hɔpene/
Infinitiv	kjøp-e	/'çø:pe/	hend-e	/'hene/
Presens	kjøp-er	/'çø:per/	hend-er	/'hener/
Preteritum	kjøp-te	/'çø:pte/	hend-te	/'hente/
Ubest. sing.	kjøp-	/'çø:p/		
Best. sing.	kjøp-et	/'çø:pe/		
Ubest. plur.	kjøp-	/'çø:p/	hend-er	/'hener/
Best. plur.	kjøp-ene	/'çø:pene/	hend-ene	/'henene/
Ubest. sing.	bøk-	/'bø:k/		
Best. sing.	bøk-en	/'bø:ken/		
Ubest. plur.	bøk-er	/'bø:ker/	bøk-er	/'bø:ker/
Best. plur.	bøk-ene	/'bø:kene/	bøk-ene	/'bø:kene/

Tabell 1. Tonelag 1 og 2 i høve til leksikalske og bøyingsmorfologiske faktorar.

Som det vil gå fram av denne tabellen, er det uråd å laga systematiske reglar for tildeling av tonem når ein ser på endingane –er, –en, –et og –ene. All erfaring med bruk av reglar for tildeling av tonem syner også at det er vanskeleg å finna eit "vanntett" regelsett. Dersom eit ord vert uttalt med feil tonem høyrer det unaturleg ut og kvaliteten på syntesen vert dårleg.

2. Tonem 1,5

Idéen om eit nøytralt tonem som kan erstatta tonem 1 og 2 i ein talesyntetisator var grunnlaget for ei undersøking utført av Björn Granström og Kjell Gustafson, presentert på konferansen "Nordic Prosody" i Odense i 1987, jfr. Granström & Gustafson (1987). Med utgangspunkt i austnorsk dialekt, vart det nøytrale tonemet, tonem 1,5, konstruert som ein mellomting av tonem 1 og 2 ved å interpolera F0-kurvane for desse som illustrert i fig. 1.



Figur 1. Tonem1,5 interpolert mellom tonem 1 og 2 for austnorsk.

Arbeidshypotesen i undersøkinga var at dersom eit automatisk system for tildeling av tonem fungerte dårleg, ville det vera betre om ein konsekvent brukte tonem 1,5 fordi dette tonemet ville bli oppfatta som enten tonem 1 eller 2 avhengig av kva kontekst ordet stod i. For å testa ut denne hypotesen vart det konstruert 2 basissetningar:

Basissetningar

- I Du må gjenta når det skal lande
- II Da tar jenta bilen ut på landet

Ut frå desse basisstningane vart det så konstruert 4 referansetningar og 3 testsetningar:

Referansetningar

- 1. Begge tonem korrekte
- 2. Første tonem feil

3. Andre tonem feil
4. Begge tonem feil

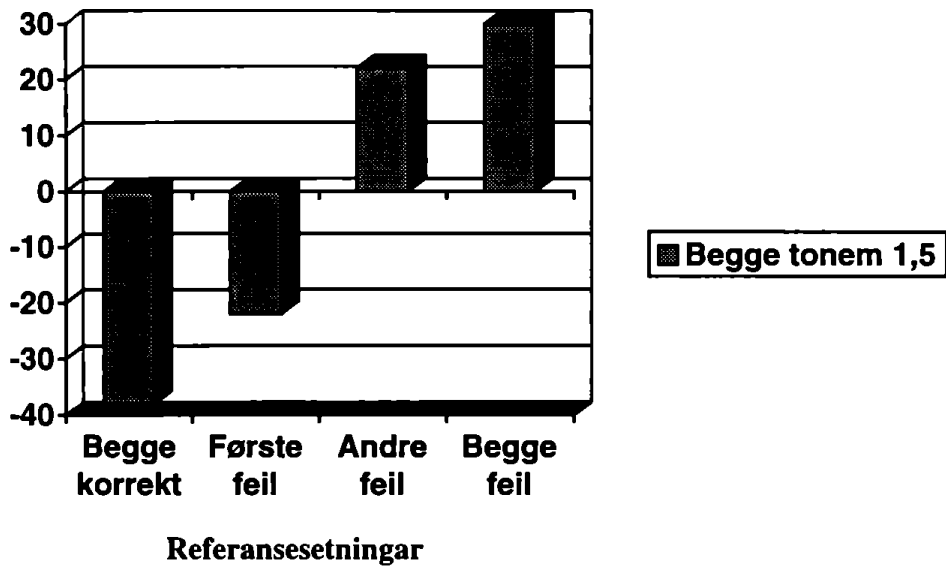
Testsetningar

1. Begge tonem 1,5
2. Første tonem korrekt, andre tonem 1,5
3. Første tonem 1,5, andre tonem korrekt

Referanse- og testsetningane vart så presenterte parvis i randomisert rekkjefølgje til ei gruppe på 14 forsøkspersonar. Dei siste 8 setningspara vart presentert ein ekstra gang i starten av testen for at forsøkspersonane skulle venna seg til lyttesituasjonen. Resultata frå desse setningane vart seinare eliminert frå testen. I eit setningspar vart første setning kalla A, og siste B. Kvar forsøksperson skulle så avgjera kva setning som høyrdest best ut langs ein 5-punkts skala:

1. A mykje betre enn B
2. A litt betre enn B
3. Begge omtrent like gode
4. B litt betre enn A
5. B mykje betre enn A

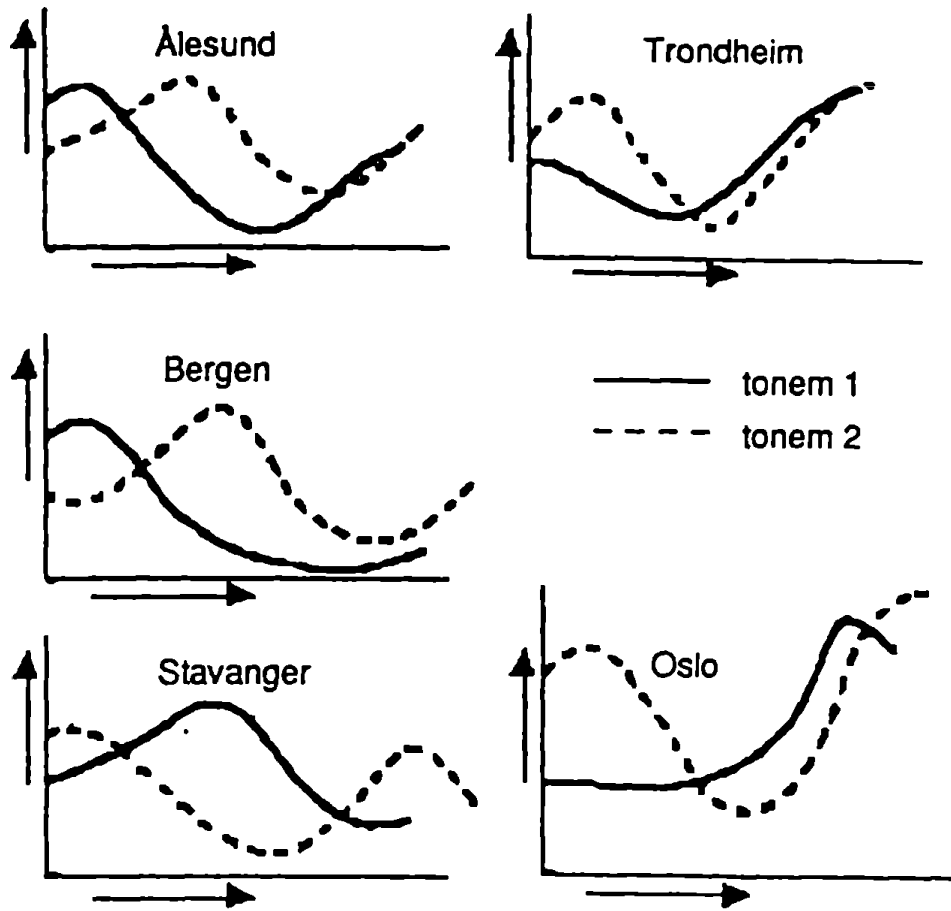
I konklusjonen for denne undersøkinga heiter det at tonem 1,5 vart vurdert som eit godt kompromiss mellom korrekt og feil bruk av ekte tonem. Dette resultatet er tufta på ein implisitt føresetnad om at resultata frå dei 3 testsetningane kan og skal vektast likt. Imidlertid lyt resultata frå denne testen tolkast på litt ulike måtar avhengig av kva applikasjon dette skal brukast i. Ein mogleg applikasjon er t.d. at ein har ei unntaksliste av høgfrekvente ord som inneheld informasjon om korrekt tonelag, og at ein berre brukar tonem 1,5 dersom ordet ikkje står i denne lista. Imidlertid vert det då vanskeleg å tolka dei resultata som gjeld testsetning 2 og 3 fordi antal tilfelle der tonem 1,5 vert brukt alltid vil vera langt større enn antal tilfelle der ein kan bruka korrekt tonelag henta frå høgfrekvenslista. Ei liste av høgfrekvente ord kan berre innehelda nokre få tusen ord, og dermed gjenspeglar ikkje resultata som gjeld testsetning 2 og 3 ein realistisk applikasjon. Dersom ein skal få fram ei meningsfull tolking av resultata frå Granström og Gustafson si undersøking, er det berre testsetningar med konsekvent bruk av tonem 1,5 som er aktuelt å vurdere resultata ut frå. Figur 2 gjev ei framstilling av resultata dersom ein samanliknar gjennomført bruk av tonem 1,5 med bruk av dei ekte tonema.



Figur 2. Vurderingar av testsetningar samanlikna med referansesetningar.
Talemålsgrunnlag: Austnorsk.

Negativ verdi på skalaen tyder altså at referansesetninga vert vurdert som betre enn testsetninga. Om ein tenkjer seg ein realistisk applikasjon der ein gitt grammatikk styrer tildeling av tonem med ein treffprosent på 50, så vil kvar av dei 4 kategoriane for referansesetningane vera like sannsynlige ($p=0,25$), d.v.s. at dei lyt vektast likt når ein skal gjera opp status. Om ein no brukar tonem 1,5 konsekvent, ser ein at resultatet vert dårlegare (negativ totalsum) enn om tonemtildelinga vert styrt av ein grammatikk som tek feil i 50% av tilfella.

Dette er eit ganske anna resultat enn det forfattarane ønskte å presentera i denne artikkelen. Det er grunn til å spørja kva grunnen til eit slikt negativt resultat kan vera. Om ein studerer tonemkonturane for ulike norske dialektar, er det rimeleg å tenkja at resultatet er avhengig av kva dialekt ein legg til grunn for undersøkinga. Figur 3 viser tonemkonturane for 5 norske dialektar.



Figur 3. Tonemkonturar for 5 norsk dialektar. Vertikal akse: F0. Horisontal akse: Tid. (Figuren er tilpassa frå Fintoft, K. 1970. *Acoustic Analysis and Perception of Tonemes in some Norwegian Dialects*. Oslo: Universitetsforlaget.)

Som det går fram av figur 3, har Oslo-dialekten nokså stor skilnad mellom F0-konturane for tonem 1 og 2. Særleg gjeld dette den første og trykksterke fasen av tonemrealiseringa som tilsvarear omtrent 1. halvpart av konturen. Det er denne fasen som er perseptorisk viktig når det gjeld å kunna skilja dei 2 tonema frå kvarandre, og av den grunn er Oslo-dialekten lite eigna som utgangspunkt for eit nøytralt, mellomliggjande tonem. Det er mange måtar å måla skilnad mellom tonemkonturar på, men 2 kriterium synest vera naturleg å festa seg ved:

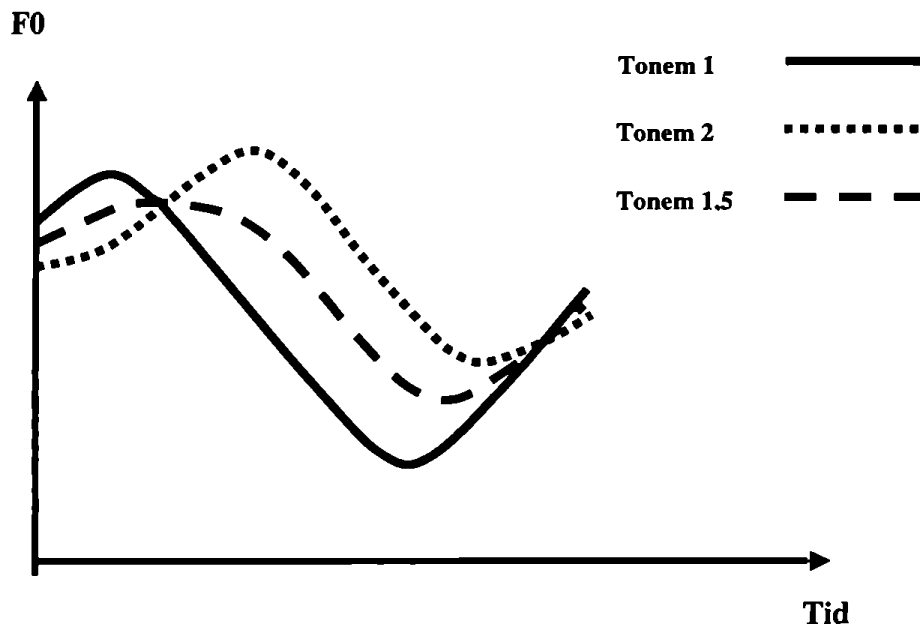
- Absoluttverdien av F0-skilnaden langs ulike tidspunkt i den 1. fasen av realiseringa
- Spektral vs. temporal konturskilnad

Målt etter begge kriteria, ser ein at Ålesunds-dialekten har ein relativt liten skilnad mellom tonemkonturane samanlikna med dei andre dialektane. Særleg tydeleg vert dette dersom ein fokuserer på den temporale skilnaden: Ved å parallellforskyva konturen for

tonem 1 langs tidsaksen, vil han falla saman med konturen for tonem 2. Dette er ikkje råd å få til med tilsvarande konturar frå nokon av dei andre dialektane.

3. Nytt forsøk

Vi ønskte å undersøkje om andre dialektar enn Oslo-dialekten kunne vera betre eigna for ei tilsvarande undersøking. Ålesunds-dialekten skilde seg naturleg ut i denne samanhengen av dei grunnar som er nemnde ovanfor. Figur 4 viser tonemkonturane for Ålesunds-dialekten der også konturen for tonem 1,5 er teikna inn.



Figur 4. Tonemkonturar for Ålesunds-dialekten.

Det går tydelig fram av figur 4 at avstanden frå konturen til tonem 1,5 og til ein av dei andre konturane er klart mindre enn tilsvarande avstand i fig. 1. Undersøkinga vart utført etter same lest som hos Granström og Gustafson, men med nokre viktige modifikasjonar:

- Fastsetjing av testintern 0-linje v.h.a. kontrollpar som var akustisk identiske
- Bruk av standard statistiske metodar (Wilcoxons signed rank test)
- Statistisk kontroll av likeverdighet m.o.t. effekten av tonem 1,5 for dei 2 basissetningane I og II
- Presentasjonsrekkefølge går ut som variabel

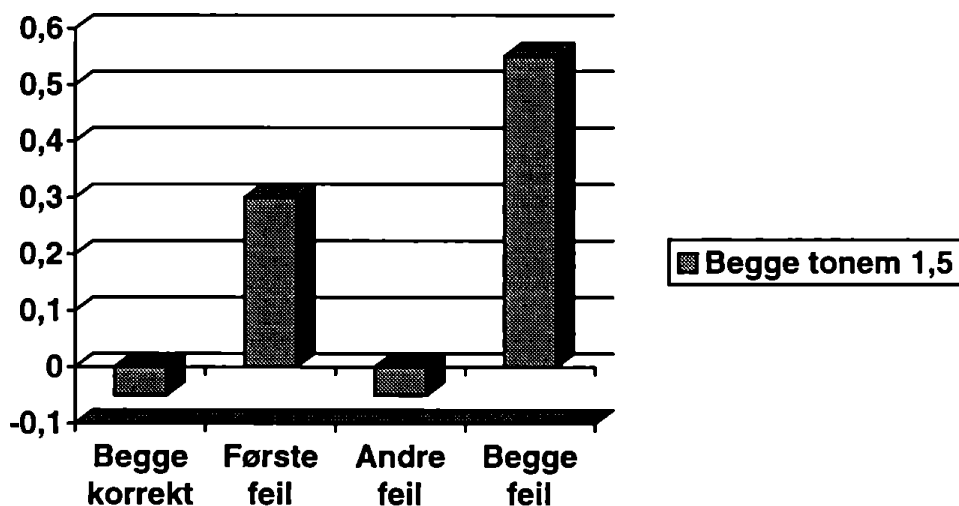
Ei gruppe på 13 forsøkspersonar vart sett saman av personar med fonetisk og lingvistisk bakgrunn, d.v.s. med både teoretisk og praktisk skulering i begge disiplinane. Alle gjennomgjekk først ein test der det vart undersøkt om dei hadde tilstrekkeleg "kompetanse" i å diskriminera tonelag på bakgrunn av ein lyttetest. Resultata frå 3 av

desse personane måtte utelatast frå testmaterialet fordi dei ikkje klarte den innleiande testen.

Fastsetjing av testintern 0-linje er viktig for å få til ei innbyrdes kalibrering av dei ulike forsøkspersonane. Det var nemleg store individuelle skilnader mellom forsøkspersonane m.o.t. korleis dei oppfatar setningar med akustisk identiske stimuli. Ved å spela av 6 setningspar som kvar inneheldt 2 akustisk identiske setningar, kunne ein registrera slike individuelt betinga avvik frå "0" som testresultat. Slike avvik må då skuldast eigenskapar hos den einskilde person og ikkje eigenskapar ved testmaterialet. Ein forsøksperson hadde 0,83 i snitt medan andre hadde -0,16 i snitt på denne testen. Det er difor viktig å kalkulera inn denne skilnaden i den vidare utrekninga elles risikerer ein at faktorar som er knytte til den einskilde forsøkspersonen øydelegg resultatet av undersøkinga.

Ei psykoakustisk undersøking som denne lyt gjennomførast med standard statistiske metodar for å vera vitenskapleg gyldig. Den største veikskapen ved undersøkinga til Granström og Gustafson er at dei dreg konklusjonar av talmaterialet utan å underkasta det vanlege statistiske metodar for hypoteseprøving. Sett frå ein vitenskapleg synsvinkel vert dermed konklusjonane ugyldige, og det er ikkje råd å seia om dei resultata dei kjem fram til skuldast faktiske skilnader eller tilfeldige samantreff. Av den grunn har vi heller ikkje gjort ei direkte samanlikning mellom resultata frå dei to undersøkingane.

Det er ikkje sjølvstyk at dei 2 basissetningane "Du må gjenta når det skal lande" og "Da tok jenta bilen ut på landet" er likeverdige m.o.t. effekten av tonem 1,5. Dette lyt også undersøkjast med statistiske metodar. Resultatet viste ingen signifikant skilnad mellom desse setningane på dette området³, og det aritmetiske gjennomsnittresultatet for desse 2 basissetningane kan difor nyttast for kvar av dei 24 moglege setningspara i testen. Figur 5 viser resultatet av undersøkinga. Berre dei resultata som tilsvarar resultata i figur 2 er tekne med.



Referansesetningar

Figur 5. Vurderingar av testsetningar samanlikna med referansesetningar.
Talemålsgrunnlag: Ålesund bymål.

Som det går fram av figur 5, kjem tonem 1,5 best ut i 2 tilfelle, medan dei ekte tonema kjem best ut i dei 2 andre tilfella. Imidlertid er ingen av desse resultatata statistisk signifikante på 5% nivå, sjølv om tendensen er klart i favør av tonem 1,5. På 10% forkastningsnivå er det siste resultatet (tonem 1,5 kontra begge tonem feil) så å seia signifikant.

4. Konklusjon

Bruk av tonem 1,5 er like bra eller betre som bruk av ekte tonem i alle setningspara der konsekvent bruk av tonem 1,5 vert samanlikna med bruk av ekte tonem. Val av dialekt er avgjerande for utfallet av ei slik undersøking.

Fotnotar

¹ Problemet med korrekt tildeling av tonem er like aktuelt for svensk talesyntese, sidan dei aller fleste svenske dialektar også er tonemiske. Imidlertid er problema litt færre i svensk sidan nestan alle kompositaord får tonem 2, mens det same ikkje er tilfelle for norsk.

² Tabellen er utarbeidd av Kjell Gustafson og presentert i artikkelen "Some innovations in a Norwegian text-to-speech system", pp. 81-85 trykt i *Quarterly Progress and Status Report*, Stockholm, 1/1989.

³ Wilcoxon signed-rank, tosidig testing. Minste verdi: $p=0,32$.

Litteratur

- Carlson, R., Granström, B. & Hunnicutt, S. 1982. A multi-language text-to-speech module. *Proceedings ICASSP- Paris*, 3,1604-1607.
- Carlson, R. & Granström, B. 1986. Linguistic processing in the KTH multi-lingual text-to-speech system, *Proceedings ICASSP- Tokyo*. 4, 2403-2406 .
- Granström, B. & Gustafson, K. 1987. Toneme 1 ½ in a Norwegian Text-to-Speech System. *Nordic Prosody IV, Papers from a symposium*, Odense: Odense University Press.