

RESULTATEN TQE-SURVEY

Laura van Eerten – TST-Centrale

Inleiding

In het kader van het CLARIN-NL-project Transcription Quality Evaluation (TQE) is een gebruikersenquête uitgevoerd onder contacten van de TST-Centrale, het Max Planck Instituut en de afdeling CLST van de Radboud Universiteit Nijmegen.

Het doel van het TQE-project is om een tool (webapplicatie) te ontwikkelen, waarin fonetische transcripties volledig automatisch op kwaliteit worden beoordeeld. De invoer van de tool is een corpus met audiobestanden en bijbehorende fonetische transcripties. De tool is in staat de audiobestanden automatisch op te lijnen met de transcripties, de foneemgrenzen te bepalen en voor elke segment-foneem-combinatie aan te geven hoe goed de match is (op een schaal van 0-100).

Om ervoor te zorgen dat de uiteindelijke tool zo goed mogelijk aansluit op de wensen van de gebruikers, is er een survey opgezet met 12 open en gesloten vragen over het gebruik van spraakbestanden, transcripties, software etc. De resultaten van deze survey worden gebruikt voor het verder ontwikkelen van de tool.

Resultaten

De enquête is door 34 respondenten ingevuld¹. Bij alle gesloten vragen waren meerdere antwoorden mogelijk. Daarnaast heeft niet iedereen alle vragen beantwoord (de vragen waren niet verplicht). De totalen verschillen daarom per vraag.

Bij de open vragen is een indeling gemaakt in de verschillende antwoorden. Bij de gesloten vragen zijn de resultaten in een grafiek weergegeven. In de conclusie zijn de percentages naar boven afgerond.

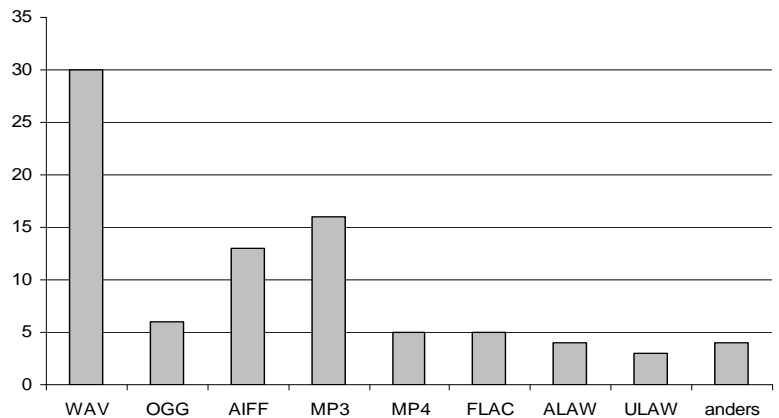
1) Vakgebied en belang van fonetische transcripties:

Antwoord	Telling	%
Spraakonderzoek algemeen (fonologie/fonetiek)	10	29,41
Onderzoek naar uitspraakvariatie	5	14,71
Forensisch spraakonderzoek	2	5,88
Taal- en spraaktechnologie algemeen (spraakherkenning/-synthese)	9	26,47
Onderzoek naar spraakontwikkeling/taalverwerving	4	11,76
Ontsluiting spraakdatabases	2	5,88
Onderzoek naar pathologische spraak	1	2,94
Documenteren van bedreigde talen	1	2,94

¹ De contactgegevens van de respondenten worden uitsluitend gebruikt om contact op te kunnen nemen met degenen die aangegeven hebben mee te willen werken aan het testen van de tool.

2a) De bestandsformaten (meerdere antwoorden mogelijk):

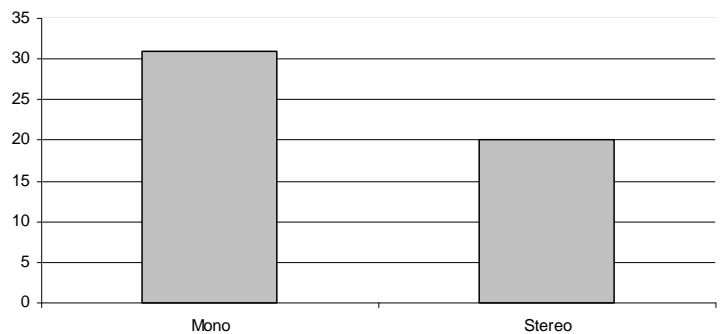
Antwoord	Telling	%
WAV	30	34,88
OGG	6	6,98
AIFF	13	15,12
MP3	16	18,60
MP4	5	5,81
FLAC	5	5,81
ALAW	4	4,65
ULAW	3	3,49
anders	4	4,65



Onder “anders” werd genoemd “raw pcm”, “wma”, “mov” en “ra”.

2b) De kanalen (meerdere antwoorden mogelijk):

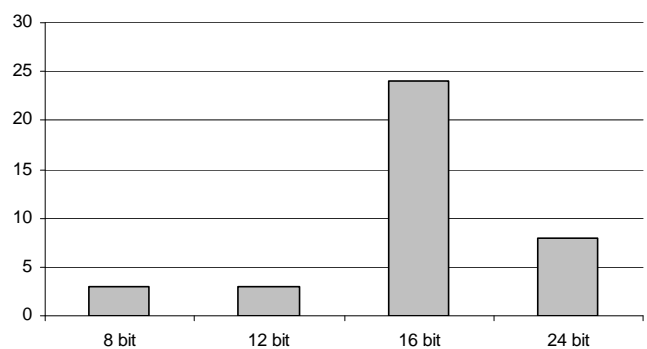
Antwoord	Telling	%
Mono	31	60,78
Stereo	20	39,22



17 respondenten hebben aangegeven zowel mono als stereo-opnamen te gebruiken.

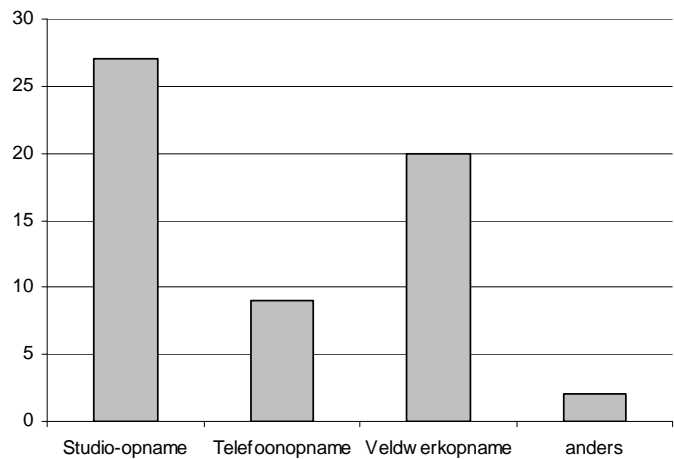
2c) De opnameprecisie (meerdere antwoorden mogelijk):

Antwoord	Telling	%
8 bit	3	7,89
12 bit	3	7,89
16 bit	24	63,16
24 bit	8	21,05



2d) De bandbreedte (meerdere antwoorden mogelijk):

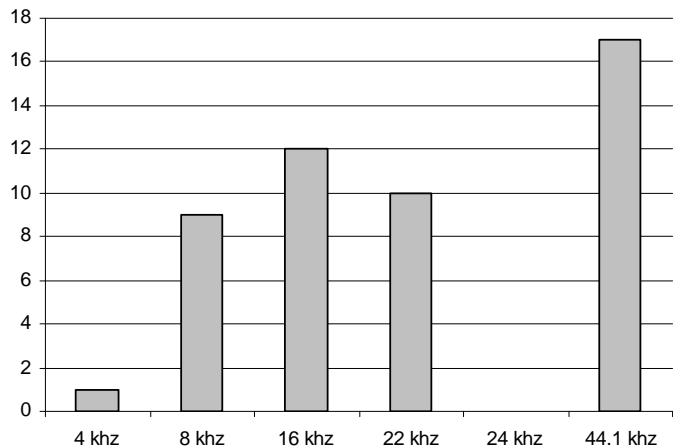
Antwoord	Telling	%
Studio-opname	27	46,55
Telefoonopname	9	15,52
Veldwerkopname	20	34,48
anders	2	3,45



Onder “anders” werd genoemd “forensisch materiaal, m.n. via telefoon” en “synthese”.

2e) De samplefrequentie (meerdere antwoorden mogelijk):

Antwoord	Telling	%
4 khz	1	2,04
8 khz	9	18,37
16 khz	12	24,49
22 khz	10	20,41
24 khz	0	0,00
44.1 khz	17	34,69



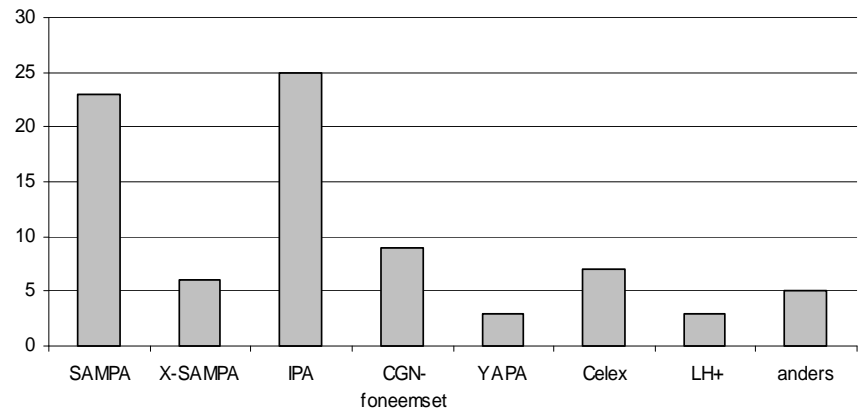
2f) Overige opmerkingen/aanvullingen bij de spraakbestanden:

Antwoord

- De antwoorden zijn van toepassing op huidig project, kan in de toekomst veranderen
- Forensisch materiaal is vaak van slechte kwaliteit; te slecht voor automatische transcripties of analyses
- In een huidig project wordt gewerkt met 8-kanaals opnamen: 8 x 44.1 khz, mono
- Het gaat om zowel audio als video waarbij de kwaliteit niet altijd even goed is
- Enkele corpora opgenomen in verschillende formats
- Gebruik gemaakt van a-law en mu-law compressie

3) De formaten en standaarden voor fonetische transcripties (meerdere antwoorden mogelijk):

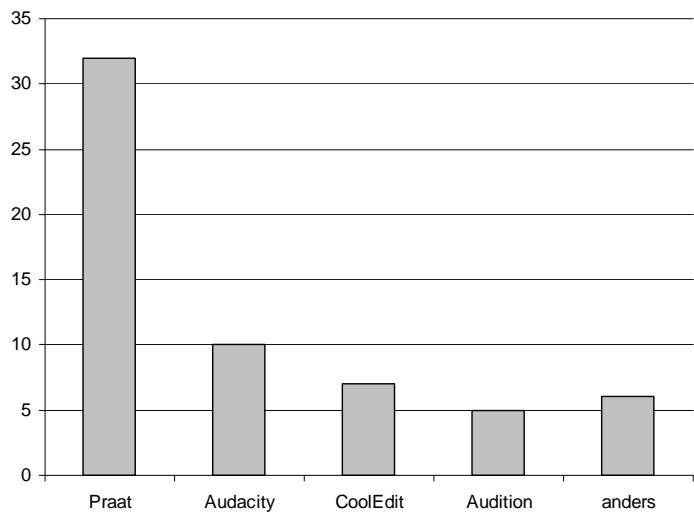
Antwoord	Telling	%
SAMPA	23	28,40
X-SAMPA	6	7,41
IPA	25	30,86
CGN- foneemset	9	11,11
YAPA	3	3,70
Celex	7	8,64
LH+	3	3,70
anders	5	6,17



Onder “anders” werd genoemd “KIPA” (2x), “Fluency-notatie (gebaseerd op SAMPA)”, “eigen codering” en “Polyphone”.

4) De software (meerdere antwoorden mogelijk):

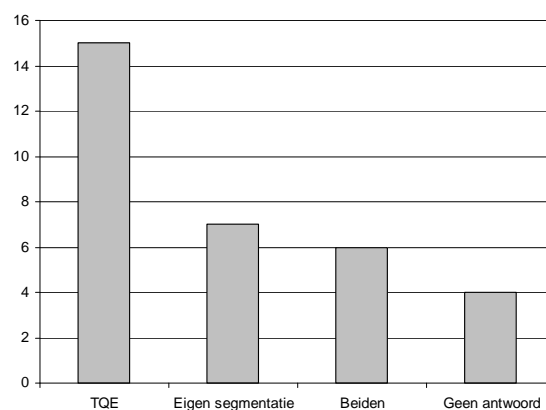
Antwoord	Telling	%
Praat	32	53,33
Audacity	10	16,67
CoolEdit	7	11,67
Audition	5	8,33
anders	6	10,00



Onder “anders” werd genoemd “eigen software (uit de SPRAAK-toolkit)”, “Elan” (2x), “Goldwave”, “Company Inhouse Tools” en “Gipos”.

5) Foneemgrenzen/segmentatie door TQE laten bepalen of eigen segmentatie meeleveren (meerdere antwoorden mogelijk):

Antwoord	Telling	%	Reden
TQE	15	46,88	<ul style="list-style-type: none"> • Ongebiasd en repliceerbaar • Sneller om op basis van TQE te hersegmenteren i.p.v. alles handmatig te doen • Hangt af van onderzoeksvraag, verbetering moet in ieder geval mogelijk zijn • Alleen als foneemgrenzen niet van direct belang zijn voor het onderzoek • Gemak • Segmentatie op foneemniveau te duur voor grote hoeveelheden spraak • Consistente data om te distribueren • Geen eigen segmentatie • TQE vindt wellicht meer of minder fonemen, de gegeven oplijnen klopt dan niet meer
Eigen segmentatie	7	21,88	<ul style="list-style-type: none"> • Semi-geautomatiseerde annotatie is nauwkeuriger • Automatische segmentatie bevat te veel fouten • Kwaliteit eigen segmentatie verifiëren door TQE • Er wordt met onbeschreven talen gewerkt • Woordgrenzen komen niet altijd overeen met prosodische uitingen van kinderen
TQE + eigen segmentatie	6	18,75	<ul style="list-style-type: none"> • Om te vergelijken: automatische segmentatiefouten eruit halen • Afhankelijk van onderzoeksvraag, doel en kwaliteit van het materiaal (goede studiokwaliteit > automatisch) • Afhankelijk van hoe data voorhanden is, manuele segmentatie heeft de voorkeur • Hangt af van de kwaliteit van het programma in het bepalen van de grenzen
Geen antwoord	4	12,50	<ul style="list-style-type: none"> • Weet niet hoe en of ik TQE ga gebruiken • Gebruik van fonetische transcripties wordt nog niet in de praktijk toegepast > audio- en tekstproductie voor spraaksynthese zijn gescheiden • Dit doet een andere afdeling (onderzoekers)



6) Gemiddelde lengte van de spraakverzamelingen:

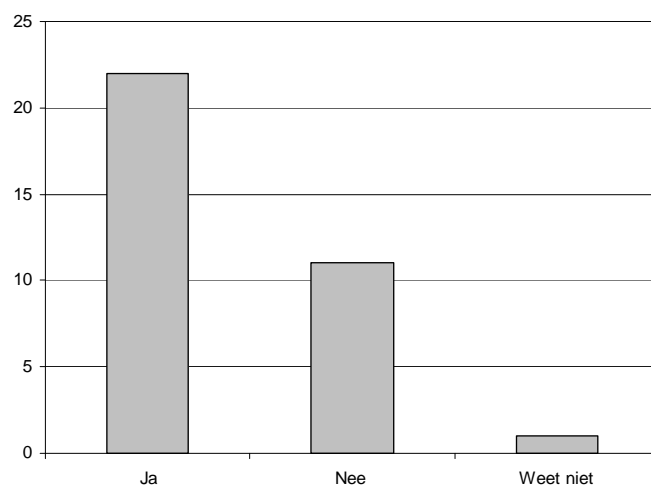
Antwoord	Telling	%
Minder dan een uur (gemiddeld 15-30 minuten per fragment/bestand)	14	48,28
Meer dan een uur (gemiddeld 1-10)	4	13,79
Tientallen tot honderd uren	3	10,34
Enkele GB's	3	10,34
Enkele honderden MB's (tot 700 MB)	4	13,79
Enkele TB's	1	3,45

7) Verwacht gebruik van TQE:

Antwoord	Telling	%	Reden
Ja	14	41,18	<ul style="list-style-type: none"> • Annoteren van corpora • Syntactisch onderzoek naar clitisatie • Transcripties spontane spraak • Controle eigen/handmatige transcripties • Hulp voor training transcibenten • Analyse prosodische- en uitspraakvarianten in L1 en L2 Nederlands en Engels • Controle G2P • Doorzoekbaar maken van interviews • Oplijnen transcriptie voor patiëntenspraak • Validatie transcripties kindertaal • Checken stimulusmateriaal
Nee	7	20,59	<ul style="list-style-type: none"> • Engelstalige data, bestanden delen via surfgroepen of dropbox • Tevreden met huidige toepassing (Phon, Childes)
Weet niet	13	38,28	<ul style="list-style-type: none"> • Nog niet voldoende informatie • Misschien voor specifieke datacollecties, maar voor forensisch materiaal waarschijnlijk niet (vertrouwelijk) • Misschien als aanvulling bij huidige alignment software • Eerst testen • Mogelijk kunnen uitzonderingen gevonden worden met vergelijking van spraaksynthese en studio-opname • Misschien in de toekomst, op dit moment nog niet • Wel benieuwd hoe goed het werkt • Eventueel voor het testen van fonemen

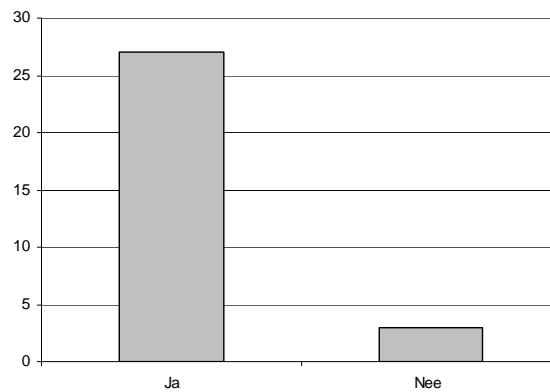
8) Interesse in opname CLARIN-infrastructuur:

Antwoord	Telling	%	Reden
Ja	22	64,71	<ul style="list-style-type: none"> • Hoe meer mensen het materiaal kunnen gebruiken hoe beter • Alleen als het gaat om niet-vertrouwelijk materiaal • Misschien problematisch m.b.t. auteursrecht • Mits aan de juiste voorwaarden is voldaan (instemming medeonderzoekers etc.) • Vooral beschikbaarstelling is vaak ondergeschoven kindje • Vergemakkelijken van samenwerkingsverbanden • Ik hoop dat deze infrastructuur het gemakkelijk maakt en niet over IPR begint • Niet in een positie om hierover te beslissen • Ook beschikbaar via Phon (Childes), bekend bij taalverwerwers
Nee	11	32,35	<ul style="list-style-type: none"> • Opslag op openbare server vergroot privacy-issues • Materiaal afkomstig van LDC/ELDA/TST-Centrale en mag niet verder verspreid worden • In contract met sprekers vastgelegd dat het materiaal niet voor andere doeleinden gebruikt wordt • Auteursrechten • Data komt van bedrijven/commercieel eigendom
Weet niet	1	2,94	<ul style="list-style-type: none"> • Onvoldoende bekend met het project



9) Bereid tot meeleveren metadata:

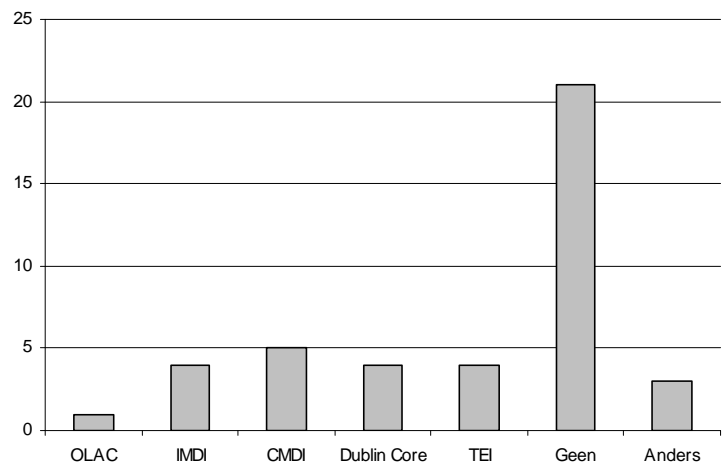
Antwoord	Telling	%
Ja	27	90
Nee	3	10



Enkele respondenten merkten op dat de privacy van de proefpersonen wel gewaarborgd moet worden.

10) Huidig gebruik van metadataforma(a)t(en):

Antwoord	Telling	%
OLAC	1	2,38
IMDI	4	9,52
CMDI	5	11,9
Dublin Core	4	9,52
TEI	4	9,52
Geen	21	50
Anders	3	7,14



Onder “anders” werd genoemd “*plain text*”, “*SAM (Eagles)*” en “*CHAT/CHILDES*”.

11) Bereid tot deelname test van TQE:

Antwoord	Telling	%
Ja	17	50
Nee	17	50

12) Overige opmerkingen TQE-project:

Antwoord

- Goed idee, hopelijk komt het van de grond
- Omschrijving nogal algemeen, moeilijk om vragen te beantwoorden
- Is het gebruik gratis of moet ervoor betaald worden?
- Nuttig project, aankondigen via o.a. NVFW en LOT
- Een transcriptietool met goede interface, ondersteuning voor verschillende alfabetten en navigatie van spraakbestanden en ingevoerde transcripties ed. zou ook mooi zijn
- Inzet van TQE binnen organisatie graag bespreken met ontwikkelaars

Conclusies

- De respondenten gebruiken fonetische transcripties voornamelijk voor diverse soorten taalkundig spraakonderzoek (29%; documenteren en analyseren) en voor toepassingen in taal- en spraaktechnologie (26%).
- Het meest gebruikte bestandsformaat onder de respondenten is WAV (37%). Daarna worden mp3 (20%) en AIFF (16%) het meest gebruikt.
- Er wordt voornamelijk gewerkt met hoge kwaliteit geluidsopnames van 16 bit (63%) en 24 bit (21%) in mono (61%).
- Onder de respondenten wordt voornamelijk gebruik gemaakt van studio-opnames (47%) en veldwerkopnames (35%).
- 44.1 khz is de meest gebruikte samplefrequentie (35%). Daarna volgt 16 khz (25%), 22 khz (20%) en 8 khz (18%). 24 khz (0%) en 4 khz (2%) wordt nauwelijks gebruikt.
- IPA (31%) en SAMPA (28%) zijn de meest gebruikte fonetische fonts.
- Praat (53%) is de software die verreweg het meest gebruikt wordt.
- Het merendeel van de respondenten (47%) zou de foneemgrenzen door TQE willen laten bepalen. Dit is vooral uit praktische overwegingen. 19% wil daarnaast ook een eigen segmentatie aanleveren om te kunnen vergelijken. De voornaamste reden dat de respondenten een eigen segmentatie aan willen leveren is omdat er weinig vertrouwen is in een automatische segmentatie.
- Het merendeel van de respondenten (48%) werkt met spraakverzamelingen van minder dan een uur.
- 41% van de respondenten verwacht TQE te gaan gebruiken voor annotatie en validatie van spraakmateriaal voor diverse doeleinden. 38% is nog niet overtuigd maar is wel bereid TQE te proberen. 21% heeft geen directe behoefte aan TQE.
- 64% van de respondenten staat positief tegenover het opnemen van eigen materiaal in de CLARIN-infrastructuur. Er zijn wel twijfels over auteursrechten. De respondenten die hebben aangegeven dit niet te willen (32%) wijten dit aan commercieel materiaal, materiaal van derden dat niet verder verspreid mag worden en auteursrechten.
- Vrijwel alle respondenten (90 %) zijn bereid om metadatabestanden aan te leveren, mits de privacy gewaarborgd blijft.
- De helft van de respondenten (50%) maakt op dit moment nog geen gebruik van (een) metadata(formaat).
- De helft van alle respondenten (17) wil deelnemen aan het testen van de TQE-applicatie.