NVFW
Nederlandse Vereniging voor Fonetische Wetenschappen

Dag van de Fonetiek 2019

Vrijdag 13 december 2019

VOC-zaal, Bushuis

Universiteit van Amsterdam

Kloveniersburgwal 48

Phonetic input under experimental control in "real" conversations: The ventriloquist paradigm

Mirjam Broersma
Centre for Language Studies, Radboud University, Nijmegen

During conversations, interlocutors adapt to each other in pronunciation and perception. A major challenge for the study of this phonetic accommodation and perceptual adaptation is the intrinsic impossibility to control the phonetic input that participants are exposed to in dialogue. The number of times participants hear certain speech sounds, the phonetic contexts in which these sounds occur, and their specific realizations, which affect the extent to which participants accommodate their pronunciation and adapt their perception to their interlocutor, thus vary across participants. The Ventriloquist paradigm (Felker, Troncoso-Ruiz, Ernestus, & Broersma, 2018) has been developed to tackle this problem. It enables the study of sound learning in dialogue, while allowing full control over the phonetic detail of the input that participants are exposed to. The Ventriloquist paradigm has been developed to investigate phonetic accommodation and perceptual learning in an ecologically valid yet maximally controlled way. Participants take part in a dialogue which they believe to be genuine; in fact, however, their real-life interlocutor is a confederate whose speech is not just 'scripted' (as in the confederate scripting task for the study of syntactic accommodation), but fully prerecorded. This guarantees control over all characteristics of the speech input, including the number of times the participant hears certain speech sounds, their phonetic contexts, and their phonetic realization. The set-up is fully tuned to upholding the illusion that the confederate is actually speaking with the participant. The confederate sits opposite the participant, face briefly hidden when (s)he "speaks". Participants hear the prerecorded speech over closed headphones. In addition to the standard input, to facilitate a smooth flow of the conversation, the confederate can play prerecorded non-verbal interaction markers and stop-gap replies to any unanticipated remarks or questions from the participant. The new paradigm thus reconciles ecological validity with experimental control for the study of phonetic accommodation in dialogue.

References

Felker, E., Troncoso-Ruiz, A., Ernestus, M., & Broersma, M. (2018) The ventriloquist paradigm: Studying speech processing in conversation with experimental control over phonetic input. *The Journal of the Acoustical Society of America*, 144, EL304-EL309. Doi: 10.1121/1.5063809.

Does short-term phonetic accommodation lead to long-term sound change? Not directly

Cesko C. Voeten
Leiden University Centre for Linguistics
Leiden Institute for Brain and Cognition

The change-by-accommodation model ('CAM'; [1,2]) suggests that sound change is caused by phonetic accommodation ([3,4]). This is tested empirically via longitudinal experiments using three on-going sound changes in Dutch: the diphthongization of /e:,ø:,o:/ ([5,6]), the blocking of diphthongs before coda /l/ ([6,7]), and the gliding of coda /r/ to [ɪ] ([8]). These have effectively completed in the Netherlands, but have not affected Flanders. This makes it possible to perform empirical studies of the CAM via *sociolinguistic migrants* ('SMs'): Flemish speakers of Dutch who migrated to the Netherlands to start their university studies.

Over the course of nine months, ten SMs and ten controls participated in three sessions of experiments focused on their production and perception of the three sound changes. The low number of participants is compensated by a high number of experimental items combined with the repeated-measures design, resulting in sufficient power. Results show robust differences between the groups, that do not convincingly diminish over time.

This is problematic for the CAM, especially because a follow-up cross-sectional experiment over multiple years' time *did* find changes. Since phonetic accommodation is known to be extremely rapid ([3,4]), the result that nine months were not enough, but decades are, casts doubt on the change-by-accommodation model.

References

[1] Auer, Peter, & Hinskens, Frans. 2005. The role of interpersonal accommodation in a theory of language change. *Dialect change: Convergence and divergence in European languages*, 335.

[2] Sonderegger, Morgan. 2012. *Phonetic and phonological dynamics on reality television*. Doctoral dissertation, Chicago.

[3] Pardo, Jennifer S. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 1194, 2382–2393. doi:10.1121/1.2178720

[4] Maye, Jessica, Aslin, Richard N, & Tanenhaus, Michael K. 2008. The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 323, 543–562. doi:10.1080/03640210802035357

[5] Van de Velde, Hans. 1996. *Variatie en verandering in het gesproken Standaard-Nederlands*. Doctoral dissertation, Nijmegen. doi:2066/146159

[6] Voeten, Cesko Cis. 2015. *The interaction between the segmental and the prosodic phonology in the midst of an on-going sound change. resolving a contradiction in the synchronic phonology of Dutch*. Master's thesis, Nijmegen. Retrieved from https://theses.ubn.ru.nl/handle/123456789/628

[7] Berns, Janine, & Jacobs, Haike M G M. 2012. A first glance at the role of length in production and perception of diphthongs before Dutch coda l. *Linguistics in the Netherlands*, 291, 15–26. doi:10.1075/avt.29.02ber

[8] Sebregts, Koen. 2015. The sociophonetics and phonology of Dutch /r/. Doctoral dissertation, Utrecht. doi:1874/306415

Is de spontane spraak van kinderen met een cochleair implantaat even verstaanbaar als die van normaalhorende leeftijdsgenoten?

Nathalie Boonen, Hanne Kloots en Steven Gillis
Centrum voor Computerlinguïstiek en Psycholinguïstiek (CLiPS), Universiteit Antwerpen

Deze bijdrage focust op de verstaanbaarheid van kinderen met een cochleair implantaat, d.i. de mate waarin "de hoorder in staat is om in de geluidsstroom woorden te herkennen, in een specifieke volgorde" (Rietveld & van Heuven, 2001, p.7). Verstaanbaarheid wordt meestal onderzocht op basis van een Likertschaal of transcripties. Bij spontane spraak liggen transcripties echter moeilijk omdat ze niet vergeleken kunnen worden met een modeltranscriptie. Een Likertschaal is meestal weinig gedetailleerd, waardoor snel maximumscores behaald worden. In deze presentatie wordt daarom een alternatieve analysetechniek voor spontane spraak toegepast: entropie.

Van zestien zevenjarige Vlaamse lagereschoolkinderen met een cochleair implantaat (CI) en zestien normaalhorende leeftijdsgenoten werden korte spontane uitingen voorgelegd aan 105 luisteraars die de uitingen orthografisch transcribeerden. De transcripties werden op woordniveau opgelijnd en geanalyseerd d.m.v. entropie. Hierbij wordt de mate van “chaos” gemeten. Principe: hoe groter de overeenkomst tussen transcripties, hoe minder “chaos” en hoe lager de entropiescore. We verwachtten dat de CI-kinderen wat minder goed verstaanbaar zouden zijn, maar ook dat hun verstaanbaarheid onderling zou variëren (Montag et al., 2014; Peng et al., 2004).

De transcripties van de uitingen van normaalhorende kinderen kwamen onderling sterker overeen dan die van CI-kinderen. Zevenjarige normaalhorende kinderen zijn dus verstaanbaarder dan leeftijdsgenoten met een CI. De scores van de CI-groep varieerden, maar driekwart behaalde scores die vergelijkbaar zijn met die van normaalhorende kinderen.

Referenties

Montag, J. L., AuBuchon, A. M., Pisoni, D. B., & Kronenberger, W. G. (2014). Speech intelligibility in deaf children after long-term cochlear implant use. *Journal of Speech, Language, and Hearing Research, 57*(6), 2332-2343. https://doi.org/10.1044/2014_JSLHR-H-14-0190

Peng, S.-C., Spencer, L. J., & Tomblin, J. B. (2004). Speech intelligibility of pediatric cochlear implant recipients with 7 years of device experience. *Journal of Speech, Language, and Hearing Research, 47*(6), 1227-1236. https://doi.org/10.1044/1092-4388(2004/092)

Rietveld, A. C. M., & van Heuven, V. J. (2001). *Algemene fonetiek*. Bussum: Coutinho.

Left-right asymmetry in tongue-palate contact during speech

Naomi Miller[1], Jo Verhoeven[1,2], Luc Daems[3], and Carlos Reyes-Aldasoro[4]
[1] Division of Language and Communication Science, Phonetics Laboratory, City, University of London, UK
[2] Computational Linguistics and Psycholinguistics Research Centre, University of Antwerp, Belgium
[3] Maxillo-Facial Surgery, ZNA Middelheim Hospital, Antwerp, Belgium
[4] Department of Electrical and Electronic Engineering, City, University of London, UK

Corresponding authors: jo.verhoeven@city.ac.uk, Naomi-Rachel.Miller@city.ac.uk

Electropalatography (EPG) is an instrumental technique for depicting tongue-palate contact during speech. Published palatograms frequently show left-right asymmetry in the amount of tongue-palate contact. This could arise due to several factors, including the speaker's anatomy, handedness, and language lateralisation, as well as asymmetries in the electropalate manufacture. This study investigated the direction of asymmetry in two groups of speakers categorised according to handedness.

Ten native speakers of English (5 left-handed, 5 right-handed) carried out various speech tasks, including sentence-reading and delivering a monologue. The sentence stimuli consisted of a phonetically balanced set of 460 semantically meaningful sentences (example: *An official deadline cannot be postponed*) taken from an online speech corpus (Wrench, 2019). The speech data were recorded while the speakers wore a custom-made electropalate with 62 electrodes. Left-right asymmetry in the EPG data was measured by calculating an index of asymmetry for each utterance. The index provided information about both the direction and size of the asymmetry.

The main finding was that the direction of asymmetry was such that there was more extensive tongue-palate contact on the left-hand side in four out of five speakers in both handedness groups. This observation is not consistent with the hypothesis that handedness may be a causative factor. The finding of greater contact on the left-hand side is entirely consistent with research on lip opening during speech: Graves et al. (1982) carried out a series of observational experiments which indicated that 150 out of 196 participants (76%) showed more lip contact on the left-hand side and greater lip separation on the right. It seems likely that the asymmetries in both of these articulatory mechanisms are caused by the same underlying principle – for example, language lateralisation, which holds that in most speakers, the left hemisphere of the brain is language-dominant (Knecht et al. 2000). The presentation will discuss several possible mechanisms that could cause greater tongue-palate contact on the left-hand side.

References

Graves, R., Goodglass, H. & Landis, T. (1982). Mouth asymmetry during spontaneous speech. Neuropsychologia 20, 371-381.

Knecht, S., Dräger, B., Deppe, M. et al. (2000). Handedness and hemispheric language dominance in healthy humans. Brain 123, 2512-2518.

Wrench, A. (2019). Mocha-Timit Multichannel Articulatory Database: English. http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html.

Prosody differs between objective and subjective causal relations in English

Na Hu, Aoju Chen, Hugo Quené, and Ted Sanders
Utrecht institute of Linguistics OTS, Utrecht University

Causality can be established either objectively or subjectively. The causality expressed in example (1) is objective, because both the consequence and the cause are real events; while the causality in example (2) is subjective, because "*Heidi is talented*" is one's opinion. In languages like Dutch, these two types of causality are expressed by specialized connectives (*omdat/want*). However, in English, they are both expressed by *because.* Hence, the types of causality are left unspecified on the lexical level (if other lexical cues are not present). This leads to the question of whether these two types of causality are distinguished by non-lexical cues, e.g. prosody. We investigated this question in both forward and backward causals using a dialogue task. Taking the Bayesian approach, we examined a wide range of prosodic features, including not only static measures such as pitch and duration, but also dynamic measures on the shape of pitch contours extracted by Functional Principal Component Analysis. The results showed that in comparison with objective causals, subjective causals were produced with higher F_0 maximum, lower F_0 minimum, longer duration, and also with distinctive contour shapes. These results indicate a trade-off between lexical and prosodic cues.

(1a) Heidi is thrilled because she won the first prize at the art festival. [backward]
(1b) Heidi won the first prize at the art festival so she is thrilled. [forward]
(2a) Heidi is talented because she won the first prize at the art festival. [backward]
(2b) Heidi won the first prize at the art festival so she is talented. [forward]

Morphological effects on the acoustics of word-final /s/

Tim Zee[1], Louis ten Bosch[1], Ingo Plag[2], and Mirjam Ernestus[1]
[1]Radboud University Nijmegen
[2]Heinrich-Heine University Düsseldorf

Previous research on English has shown that final /s/s in monomorphemic words are acoustically longer than /s/ suffixes, suggesting that morphology influences the articulation of segments (Plag, Homann, & Kunter, 2017). The present study extends this line of research to Dutch by investigating the duration and spectral centre of gravity (CoG) of non-suffixal /s/ (e.g., kies) and plural /s/ (e.g., ski's) across both scripted and spontaneous speech registers in Dutch speech corpora.

Models of the residualised measures showed significant interactions between register and morphological status for both duration and CoG. In conversational speech, non-suffixal /s/ was longer and had a higher CoG than plural /s/. In news broadcasts, only a durational effect was found, whereas read-aloud stories showed no morpho-acoustic effects whatsoever. These results replicate previous durational findings for English. Moreover, the additional spectral difference in conversational speech seems to reflect a general phonetic reduction of /s/ in plurals. However, the differences do not hold across scripted speech registers, suggesting a role for speech planning.

References

Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final s in English. Journal of Linguistics, 53(1), 181–216.

Testing a conflict-based theory of self-monitoring for speech errors

Sieb Nooteboom and Hugo Quené
Utrecht institute of Linguistics OTS, Utrecht University

This paper reports an experimental test of a theory of self-monitoring proposed by Nozari, Dell and Schwartz (2011). The theory presupposes that multiple items generated by the speech production system may be active simultaneously, correct and incorrect items competing for the same slot. In case of error, conflict information is passed on to an executive control center, leading to detection. When the overall conflict in the system increases, distinguishing between correct and error trial becomes more difficult, which in turn may result in (1) more errors being made, and (2) fewer errors being detected. We test this theory with data obtained in two experiments eliciting segmental speech errors in Dutch CVC CVC word pairs, as reported in Nooteboom & Quené (accepted for publication). The two experiments accidentally differed in overall conflict. All of the stimuli in the first experiment, with least overall conflict, were also used in the second experiment. There overall conflict was much higher because of the inclusion of a condition eliciting errors against the relatively weak voiced-voiceless feature in initial stop consonants, leading to more errors against this feature, and to fewer errors being detected. The crucial comparison is between the conditions with identical stimuli. There we find significantly fewer errors being detected in the second than in the first experiment. This supports a conflict-based theory of speech error detection.

References

Nooteboom S.G. & Quené. H. (accepted for publication). Repairing segmental speech errors. Competition as a source of repairs. *Journal of Memory and Language*.
Nozari, N., Dell, G., & Schwartz, M. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology* 63, 1–33.

Regional variation in plosive realization in Danish

Rasmus Puggaard, Leiden University Centre for Linguistics

It is well-known among native speakers of Danish that there is regional variation in the realization of /t/ (which has long voicing lag and is highly affricated in the standard variety). In particular, speakers of the Northern Jutlandic variety has a /t/-variant that is well-known for having relatively short voice onset time (VOT) and – as opposed to Standard Danish – no affrication. Using a large corpus of traditional dialect speakers from 213 parishes spread across the Jutland peninsula and more than 17,000 segmented plosives, it is shown that such variation in stop realization is not limited to /t/, is not limited to Northern Jutland, and is not categorical. In fact, there is continuous variation across the peninsula in VOT of aspirated stops, as well as affrication. Furthermore, affrication cannot be predicted from VOT. Generalized additive mixed models are used to statistically model the geographical variation found in the data directly without resorting to e.g. normalization by region. These models show that patterns of variation found in the data do not cluster with traditionally defined dialect regions. The variation that can be attributed to linear predictors (phonetic or otherwise) is generally in line with existing literature on the topic.

How consistently do speakers apply the Lombard speech clarification effect over time?

Chen Shen and Esther Janse {c.shen; e.janse}@let.ru.nl
Centre for Language Studies, Radboud University, Nijmegen, the Netherlands

We investigated the acoustic-phonetic differences between speakers' habitual speaking style and their speaking style in a condition where they were presented with loud noise and were also instructed to speak clearly (Lombard/clear style). Our research question was whether acoustic differences in articulation rate, pitch median, pitch range, and spectral tilt between habitual and Lombard speaking style would change over the course of a sentence list.

Seventy-eight participants read out 48 sentences (order randomised) in both their habitual style, and in Lombard style. Results from linear mixed-effects models indicate that trial main effects were present in three of the four acoustic measures (i.e., articulation rate, pitch median, and spectral tilt). Across all four acoustic measures, sentence trial interacted with speaking style. More specifically, acoustic differences between habitual and Lombard speech increased over trials, which was sometimes due to speakers becoming 'sloppy' in their habitual style over trials, e.g., faster articulation rate and smaller pitch range. However, speakers also enhanced some of their Lombard style modifications over Lombard trials, e.g., higher pitch median and flatter spectral tilt. Thus, despite the higher vocal effort, speakers in our study were able to not only maintain but even enhance their Lombard speech modifications over trials.

The role of position in native and non-native filled pause acoustics

Meike de Boer and Willemijn Heeren
Leiden University Centre for Linguistics, Leiden University

In this presentation, we expand our analysis as presented last year on native and non-native filled pauses, now considering the role of position in the utterance. Last year, based on claims by [1, 2] that non-native (L2) speakers would transfer their filled pauses directly from their native language (L1), we compared filled pauses in L1 Dutch and L2 English of female speakers, using linear mixed-effects models. We found that L2 speakers do not simply transfer their filled pauses from their L1 but change their formant realizations and the relative occurrences of *um* and *uh*.

Prior studies showed that filled pauses in different positions of the utterance might have different characteristics [e.g. 3, 4]. To test whether the language effects we found could actually be (partly) explained by position effects, we included Position in the fixed parts of our models. We will show that while Position indeed affects filled pause realization, our cross-linguistic findings remain. Overall, results show that when analyzing filled pauses, contextual factors should be considered, including their language, form (*uh* or *um*), and position.

References

[1] Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*(1), 73-111.
[2] De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics, 19*(2), 85-114.
[3] Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *Journal of Speech, Language and the Law, 23*(1), 99-132.
[4] Shriberg, E. E., & Lickley, R. J. (1993). Intonation of clause-internal filled pauses. *Phonetica, 50*(3), 172-179.

Auditory and Visual Cues in the Production and Perception of Mandarin Tones

Yueqiao Han[1], Thiago Castro Ferreira[2], Martijn Goudbeek[1], Maria Mos[1], and Marc Swerts[1]
[1]Tilburg University, [2]Federal University of Minas Gerais, UFMG

Which acoustic and visual cues can be used to classify Mandarin tones? Furthermore, are these cues the same for producers (i.e., the ground truth) and (tone-naïve) perceivers? To address these questions we video-taped four Mandarin speakers while they produced ten syllables with four Mandarin tones, i.e. forty words in two styles (natural and teaching), totaling 160 stimuli. The audiovisual stimuli were subsequently presented to 43 tone-naïve participants in a tone identification task. Basic acoustic and visual features were extracted. We used decision trees and machine learning to identify the most important acoustic and visual features for classifying the tones. These features were identified separately for the tones as produced by the four speakers and for the tones as identified by the 43 perceivers. Based on previous work, we expected that acoustic features would be more relevant than visual features, but that non-native perceivers might still benefit from the visual signal. The results showed that acoustic features were ranked higher than the visual features for tone classification, both for the classification of the intended and the perceived tone. However, tone perceivers did revert to the use of visual information in certain cases. So, while visual information does not seem to play a significant role in native speakers' tone production, tone-naïve perceivers do sometimes consider visual information in their tone identification.

Self-interruptions and repetitions in spontaneous dialogues and their speaker-specificity

Sanne Ditewig, Utrecht University

Disfluencies, such as self-interruptions and repetitions, provide a window on the speech production and self-repair process. Most knowledge on these disfluency types comes from experimental tasks and monologues [e.g. 1,2]. However, most of everyday speech consists of spontaneous dialogues, a speech style which can be argued to differ markedly from experimentally elicited speech, because it requires more planning and aspects of dialogue management [e.g. 3,4]. In addition, little is known about inter-speaker variation in the production of disfluencies. As a result, theories on speech production and self-repair generally do not include variation. There is thus a gap in knowledge, leading to the following questions: (a) What is the distribution of self-interruptions and repetitions and their characteristics in spontaneous face-to-face dialogues and (b) how speaker-specific are they?

These questions were investigated by annotating interruptions and self-repetitions and several of their characteristics in spontaneous Dutch face-to-face dialogues and analyzing their distributions, which were compared to those obtained in earlier work on other speech styles. Specific distributions for spontaneous speech were found, suggesting that disfluency behavior depends on speech style and allowing for an evaluation of several self-repair models and theories. Moreover, self-interruptions and repetitions showed inter-speaker variation and contained some speaker-specific information.

References

[1] Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14*(1), 41-104.
[2] Levelt, W. (1984). Spontaneous self-repairs in speech: Processes and representations. In *10th International Congress of Phonetic Sciences* (pp. 105-117). Foris.

[3] Sacks, H., Schegloff, I., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50,* 696-735.
[4] Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society, 29*(1), 1-63.

Trill Type and Articulatory Setting: an EPG study

Jo Verhoeven[1,2], Naomi Miller[1], Luc Daems[3], Hanne Kloots[2], and Carlos Reyes-Aldasoro[4]
[1] Division of Language and Communication Science, Phonetics Laboratory, City, University of London, UK
[2] Computational Linguistics and Psycholinguistics Research Centre, University of Antwerp, Belgium
[3] Maxillo-Facial Surgery, ZNA Middelheim Hospital, Antwerp, Belgium
[4] Department of Electrical and Electronic Engineering, City, University of London, UK
Corresponding authors: jo.verhoeven@city.ac.uk, Naomi-Rachel.Miller@city.ac.uk

A long-standing practice in Flemish drama schools has been to promote a tongue-tip trill realisation of the phoneme /r/. Thus, actors with a natural uvular trill have been required to learn to pronounce an alveolar trill. The rationale behind this, reportedly, is the assumption that the use of an alveolar trill triggers a shift in the articulatory setting towards a more anterior position, hence improving the actor's intelligibility. It is not clear what scientific principle this expectation is based on, but it is commonly mentioned in pronunciation manuals; see, for example, Eldar (1906:159), Oostveen (1936:35), van Amelsvoort & Franssen (1960:65), Linthorst et al. (1968:64), Timmermans (2008:179), and Lacroix (2009:178).

Electropalatography data were collected from a professional speaker of Belgian Dutch who uses a naturally-acquired uvular-r in everyday speech and a learnt tongue-tip trill in her professional work. Sentences in which either the first or last word began with /r/ were firstly read with an alveolar-r, and then all sentences were re-read with a uvular-r. For each realisation, the palatograms associated with /r/ and its five neighbouring phonemes were excluded, to eliminate the effects of coarticulation. A front-back centre-of-gravity (CoG) measure was calculated from the remaining palatograms.

When /r/ occurred in sentence-final position, there was a significant ($p<0.01$, Mann-Whitney U-test) effect of trill type (median CoG 0.422 and 0.413 for alveolar and uvular, respectively). This suggests that the articulatory setting is slightly more anterior in utterances with alveolar trills. It is unlikely, however, that this small difference has a significant effect on speaker intelligibility. There was no difference in CoG when /r/ occurred in sentence-initial position.

References

F. van Amelsvoort & A. Franssen (1960). Stem en spraak. s.l., Parcival, 3th pr.

A.M. Eldar (1906). Spreken en zingen. Tiel, D. Mijs, 10th ed.

C. Lacroix (2009). Goed gestemd. Tielt, Lannoo.

P. Linthorst, G. Leerkamp & M. Galle (1968). Het spreekonderwijs op de kweekschool. Groningen/Leuven, Wolters-Noordhoff/J.B. Wolters, 10th pr.

J. Oostveen (1936). Theorie en praktijk van het spreken. The Hague, G. Naeff.

B. Timmermans (2008). Klink klaar. Uitspraak- en intonatiegids voor het Nederlands. Leuven, Davidsfonds, rev. ed.

Towards pseudonymized speech

Rob van Son
Dept. Head&Neck Oncology and Surgery, Netherlands Cancer Institute Amsterdam
ACLC, University of Amsterdam

Sharing speech recordings and speech data is important for progress in speech science and technology. However, sharing speech, whether for demonstration or for research, raises privacy concerns. There are many situations where we would like to be able to hide the identity of the speaker while still have "natural" sounding speech that exhibits all the linguistic and para-linguistic features of speech, i.e., pseudonymous speech. Current research in this direction tries to chain a speech recognizer and a speech synthesizer, e.g., a phone recognizer and a phone synthesizer, which removes the (para-)linguistically interesting aspects of the original speech. Here we explore a different approach with a continuous spectral transformation of speech using standard Praat functions, e.g., Change Gender. This results in intelligible speech varying in quality from near natural to clearly distorted. Identification in forced choice experiments is around 70% correct, both by (four) experts (range 62-75%, chance is 50%) and naive listeners (range 50-75%), indicating a loss of around 88% (range 81-99%) of the information needed for speaker identification. A final tool-box should allow users to make a trade-off between the level of pseudonymization and the preservation of relevant (para-)linguistic features. An English version of the ABX listening experiment is available from:
http://www.fon.hum.uva.nl/rob/PseudonymizedSpeechExp/

Perception of American English pure vowels by monolingual and bilingual Iranians

Vincent J. van Heuven and Naeimeh Afshar
Dept. Hungarian and Applied Linguistics, Pannon Egyetem, Veszpém, Hungary

We present preliminary results of a study on the perceptual representation of the vowel system of (American) English of monolingual learners with Persian and bilingual learners with Persian and Azeri (a Turkic language) as their mother tongue(s). Such perception studies may be used to predict and explain specific problems in the pronunciation of the target language. The first experiment explores the mapping of the 11 English vowels onto the 6 Persian) or 9 Azeri vowels of the learners within the perceptual assimilation framework (PAM). The second experiment maps out the mental conception of the English vowels in terms of vowel quality and duration (and the possible interaction between the two) of the two leaner groups. For this second experiment we developed a set of 43 reference vowel sounds synthesized in a /m_f/ context at perceptually equal distances of 1 Bark along the height (F1, 7 steps) and backness/roundness (F2, 9 steps) dimensions (excluding 20 impossible combinations) synthesized with 100 and 200-ms steady-state durations. The same 20 monolingual and 20 bilingual Iranian learners of English participated in both experiments. We test the hypotheses that (i) the bilinguals will have more narrowly defined English vowels while (ii) both groups will fail to exhibit the vowel quality-by-duration interaction that is characteristic of native English listeners.

Perception-production relationships in weighting phonetic cues of vowel contrasts

Hayo Terband[1] and Tom Lentz[2]
[1.] Utrecht University; [2.] University of Amsterdam

The present study set out to investigate phonetic cue weighting for phonological contrasts in production and perception. Specifically, we investigated the effect of normalization for individual perceptual acuity on the correlations between perceptual cue weighting and produced contrast. Our hypothesis was that production and perception are balanced based on just noticeable difference (JND) units.

Participants were 47 young adult speakers of Dutch (age 19-29). The stimuli consisted of the Dutch /ɑ/-/a/ vowel contrast (which differs both in spectral properties and duration). Measurements comprised JND's for spectral and durational differences, perceptual cue-weighting, and produced differences in formants and duration. Cue-weighting and produced contrast were converted into a non-normalized- (value_formants/value_formants +value_duration) and a JND-normalized ratio ((value_formants/JND_formants)/ (value_formants/JND_formants)+(value_duration/JND_duration)).

The results showed a significant effect of normalization and a pattern of negative correlations (if a cue is more important in perception, it is expressed less in production) turned into a pattern of positive correlations after JND-normalization. These findings suggest that production and perception are balanced based on relative perceptual acuity. The driving mechanism thus could be summarized as egocentrism or subjective balance. If speakers are perceptually more sensitive to changes on a dimension, they express objectively smaller differences on that dimension.

Agenda Algemene Ledenvergadering der
Nederlandse Vereniging voor Fonetische Wetenschappen

1. Opening
2. Mededelingen
3. Financiën: De balans over 2018 zal voor inzage beschikbaar zijn.
4. Bestuurssamenstelling
5. W.v.t.t.k.
6. Sluiting

Programma

	9:30	Inloop met koffie/thee
Keynote	10:00	Opening door de voorzitter
	10:05	Phonetic input under experimental control in “real” conversations: The ventriloquist paradigm *Mirjam Broersma*
	10:45	Pauze met koffie/thee
Oral	11:00	Does short-term phonetic accommodation lead to long-term sound change? Not directly. *Cesko Voeten*
	11:20	Is de spontane spraak van kinderen met een cochleair implantaat even verstaanbaar als die van normaalhorende leeftijdsgenoten? *Nathalie Boonen, Hanne Kloots & Steven Gillis*
	11:40	Left-right asymmetry in tongue-palate contact during speech *Naomi Miller, Jo Verhoeven, Luc Daems & Carlos Reyes-Aldasoro*
	12:00	Algemene Ledenvergadering
	12:15	Lunch op eigen gelegenheid
Oral	14:00	Prosody differs between objective and subjective causal relations in English *Na Hu, Aoju Chen, Hugo Quené & Ted Sanders*
	14:20	Morphological effects on the acoustics of word-final /s/ *Tim Zee, Louis ten Bosch, Ingo Plag & Mirjam Ernestus*
	14:40	Testing a conflict-based theory of self-monitoring for speech errors *Sieb Nooteboom & Hugo Quené*
Posters met koffie/thee	15:00 - 16:15	Regional variation in plosive realization in Danish, *Rasmus Puggard* How consistently do speakers apply the Lombard speech clarification effect over time? *Chen Shen & Esther Janse* The role of position in native and non-native filled pause acoustics *Meike de Boer & Willemijn Heeren* Auditory and Visual Cues in the Production and Perception of Mandarin Tones *Yueqiao Han, Thiago Castro Ferreira, Martijn Goudbeek, Maria Mos, & Marc Swerts* Self-interruptions and repetitions in spontaneous dialogues and their speaker-specificity, *Sanne Ditewig* Trill Type and Articulatory Setting: an EPG study, *Jo Verhoeven, Naomi Miller, Luc Daems, Hanne Kloots, & Carlos Reyes-Aldasoro*
Oral	16:15	Towards pseudonymized speech *Rob van Son*
	16:35	Perception of American English pure vowels by monolingual and bilingual Iranians, *Vincent J. van Heuven & Naeimeh Afshar*
	16:55	Perception-production relationships in weighting phonetic cues of vowel contrasts, *Hayo Terband & Tom Lentz*
	17:15	Borrel