Cognitive Load in Voice Therapy Carry-over Exercises

Iwarsson, Jenny; Morris, David Jackson; Balling, Laura Winther

Cognitive Load in Voice Therapy Carry-over Exercises

Jenny Iwarsson, David Jackson Morris, and Laura Winther Balling

Journal article (Accepted manuscript)

Cognitive Load in Voice Therapy Carry-Over Exercises

Jenny Iwarsson*, David Jackson Morris* and Laura Winther Balling**

*Speech Pathology and Audiology,
Department of Nordic Studies and Linguistics,
Copenhagen University

**Department of International Business Communication,
Copenhagen Business School

Abstract

Purpose: The cognitive load generated by online speech production may vary with the nature of the speech task. This paper examines three speech tasks used in voice therapy, in which a patient is required to adopt and automatize new voice behaviors, ultimately in daily spontaneous communication.

Method: Twelve participants produced speech in three conditions: rote speech (weekdays), sentences in a set form and semi-spontaneous speech. Subjects simultaneously performed a secondary visual discrimination task for which response times were measured. On completion of each speech task, subjects rated their experience on a questionnaire.

Results: Response times from the secondary task were found to be shortest for the rote speech, longer for the semi-spontaneous speech and longest for the sentences within the set framework. Principal components derived from the subjective ratings were found to be linked to response times on the secondary visual task. Acoustic measures reflecting fundamental frequency distribution and vocal fold compression varied across the speech tasks.

Conclusions: The results indicate that consideration should be given to the selection of speech tasks during the process leading to automation of revised speech behavior, and that self-reports may be a reliable index of cognitive load.

Background

Cognitive load is a construct representing the cognitive effort or attention required to perform a task (Kahneman, 1973; Posner, 1982). Cognitive load theory has amongst other areas been used in educational psychology, to explain how effective learning could be facilitated through the design of appropriate tasks and instructions (Sweller, 1988). Central to this concept, also called mental effort, is that attention is brought into play during the performance of a task and that attention is thought to be a limited-capacity mental resource.

Cognitive load is relevant to pedagogical strategies in behavioral motor learning and more specifically, for our interest, in voice and speech therapy. An important issue for evaluating such interventions is the carry-over requirement that a new speech behavior is successfully adopted in the patient's spontaneous everyday communication. In order to facilitate the goal of automaticity, the speech language pathologist may select exercises requiring successively increasing amounts of cognitive load so that the new speaking strategy finally becomes robust in the face of cognitive and communicative distractors. Based on clinical experience, the use of rote speech (using mechanical phrases or repetition by heart, e.g. reciting the days of the week or the months of the year) is well suited to the initial phase of training a new skill or behavior, e.g. a new respiratory strategy, a modified body posture or a more relaxed neck musculature during speech. The rationale is that rote speech involves a minimum of cognitive and communicative load and thus allows more attention to be devoted to sensory-perceptual aspects of speech production, as compared to non-rote speech. Such attention is critical in the initial phase of modifying behavior (Iwarsson, 2014). A speech task with a minimal cognitive load can be assumed to increase the potential of realizing the 'each time–every time principle', meaning that a new behavior should preferably be fired consistently, each and every time a stimulus is presented, in order to become stable and ultimately, through repetition, automatized (Verdolini & Lee, 2004). Exercises that require more focus on the content of speech (e.g. narration of the storyline of a book or explanation of how to prepare a dish) may require a greater amount of cognitive load and it may therefore be preferable to introduce these types of tasks later in the process when, by virtue of repetition, the new voice task has achieved greater automaticity. It is assumed that the higher the level of cognitive load involved in a speech task, the more it will distract the patient from diverting attention to speech and voice motor control, which is a critical aspect before a new skill or behavior is generalized to everyday speech.

Various methods have been used to measure the cognitive load of a task. *Physiological methods* include measuring pupil dilation (Kahneman & Beatty 1966, Van Gerven et al., 2004), heart rate variability (Paas et al., 1994), skin conductance (Shi et al. 2007) and electroencephalographic measures -EEG (Antonenko et al., 2010). *Behavioral methods* are based on the assumption that the cognitive load of a given task results in measurable behavioral changes, such as finger pressure on a mouse (Ikehara & Crosby, 2005), pen-input patterns (Taib et al., 2007), eye movements (Chen et al., 2011) and speech behavior (Müller et al. 2001). *Subjective ratings* offer the possibility of tapping self-assessed dimensions of a patients' mental state during the performance of a task which can reflect the level of cognitive load (Rubio et al., 2004; Luximon & Goonetilleke, 2001). Finally, there are performance-based methods to measure cognitive load.

These can be divided into two sub-categories: *primary task performance measures*, meaning that cognitive load is measured through the level of performance on the task itself, and *secondary task performance measures*, meaning that the subjects are asked to perform an additional task while simultaneously performing the main task. The secondary task is typically a cognitively simple task, such as repeating a number or responding to a signal by pressing a button. Performance on the secondary task, e.g. reaction times or number of correct responses, can then be assumed to reflect the cognitive load requirement for the main task, e.g. a speech activity. Secondary task performance measures have the advantage that the cognitive load involved in different primary tasks can be compared using the same secondary task (Posner, 1982; Paas et al. 2003; Yap, 2012).

The primary aim of the present investigation was to compare the level of cognitive load of three different speech tasks commonly used in clinical voice therapy, ultimately to facilitate the adoption of new speaking behaviors into everyday spontaneous speech. It was hypothesized that the different levels of difficulty represented in the speaking conditions would result in response time differences on a secondary task, which was visual discrimination. Another aim was to investigate the extent to which subject's self-reported ratings of cognitive load reflected their actual performance as measured by response times to the secondary task. This could advance knowledge about the reliability of a self-rating questionnaire as well as the possibility to reliably assess the level of cognitive load from subjective ratings. In addition, all speech signals were acoustically analyzed with regard to mean pitch, variation in pitch and glottal compression. This was done to document with objective measures any differences between the speech tasks, with regard to speech behaviors they induce.

Method
<u>Subjects</u>
Twelve healthy female university students, native Danish speakers, with an average age of 23 years (range 19-39 years) served as subjects. They were not informed about the aim of the study and none of them had any experience with clinical voice therapy or Japanese Kanji orthography, which was used in the secondary task. Four of the subjects were choristers and had had singing lessons, and one of these had acted professionally parallel to her her university studies. One subject had symptoms of a cold at the time of participation and another was diagnosed with Vocal Cord Dysfunction but without any symptoms at the time of participation. None of the subjects had a history of smoking, allergy, asthma or visual impairment.

<u>Speech tasks</u>
All subjects performed three speech tasks (A, B and C). Task A involved rote speech for which subjects were instructed to repeat the days of the week over and over again. In task B subjects were instructed to construct sentences in a set form that contained an element of free choice. The sentences consisted of a weekday + 'I will' + an everyday activity (e.g. 'on Monday I will ride my bike, on Tuesday I will meet a friend' and so on). Task C involved semi-spontaneous speech and subjects were instructed to describe how they would prepare a pasta sauce. The experimenter instructed the subjects to provide a detailed description of the cooking procedure,

choice of ingredients and spicing. Task instructions were given orally by the experimenter before the performance of each task. As the experimenter was not blind to the experimental hypothesis and instructions could be a factor of relevance to performance, the oral instructions were written down and kept uniform for all subjects and tasks. For all three speech tasks, the subjects were explicitly instructed not to aim at a specific tempo or meter but to just follow their own rhythm and to breathe when they needed. Also, the subjects were instructed to try each speech task once before the recording in order to be confident with the task. The order of the tasks was counter-balanced across the subjects. All instructions and speech tasks were provided and performed in Danish.

Secondary task

The secondary task was a 3-alternative forced choice visual discrimination task in which subjects were presented with 3 Kanji ideograms, of which 2 were the same and 1 was different (see Figure 1). Subjects were instructed to select as quickly and as accurately as they could the symbol that was different by pressing the corresponding button on a response box. When the button was pressed the stimuli for the next trial was presented with no inter-trial interval. Thirty randomized presentations were carried out during each speech task, making a total of 1080 responses (30 presentations X 3 speech tasks X 12 participants). Before the test, each subject completed a training round with 12 trials during which feedback on correct and incorrect responses was provided. No feedback was given during the test.

The secondary task design was based on the assumption that a long response time would reflect a high degree of cognitive load in the primary task as compared to a short response time (Kahneman, 1973; Posner, 1982). The question of single versus multiple pools of attention has been debated (Moss, 1997). Our motivation for the choice of a visual discrimination task was that we were not interested in a secondary task requiring processing in the same modality as the vocal task. In order to minimize the risk that the visual task would take the subjects' full attention, they were told that the speech task was the main task and the visual task was secondary. Also, the subjects were directed by the experimenter to begin the visual discrimination task 5 seconds after the speaking had been established. The microphone and recording were obvious to the subjects and the experimenter, who was in the room, also served as a silent communication partner in order to make the speech tasks more natural (Vogel et al. 2014).

Figure 1 about here

Subjective ratings

After each of the three speech conditions (with simultaneous visual task), the subjects completed an 8-item questionnaire that probed how they experienced the performance. Ratings were recorded on 5-point equally spaced scales (see Appendix 1 for an English translation of the questionnaire). The choice of items was motivated to a large extent by the items included in the Subjective Workload Assessment Technique (SWAT) and NASA-Task Load Index (Rubio, Dáz, Martín & Puente, 2004; Luximon & Goonetilleke, 2001), but adapted to match the requirements of measuring cognitive load in speech. Questions 1-6 were designed to probe the subject's experience of the speech task and questions 7-8 to reflect the speech and visual tasks.

Instrumentation

The visual stimuli were presented on a computer monitor (ThinkPad, Lenovo W530) and a Response Box was connected to the serial port of this computer with three active response buttons. Accuracy and response time, measured from the onset of visual presentation to the response, were recorded with a test script created and run in E-prime v 2. The audio signal from the speech tasks was recorded using a Sennheiser microphone (MKH 40 P48) placed at 40 cm from the speaker's mouth. The microphone was connected to a Behringer Ultragain High definition microphone preamp/line driver/DI-Box (Model Mic 2000), a HHB Compact Disc Recorder (CDR-850 Professional) and a Sony Integrated stereo amplifier (TA-FE510R). All recordings were made in a sound treated recording both. In addition, glottal activity and vertical larynx position were registered using a Glottal Enterprises Two Channel Electroglottograph (EG2-PCX) with two electrodes attached to the subjects' neck with an elastic ribbon and a M80 headset microphone. However, these signals were not analyzed in the present investigation.

Acoustic analysis

Five acoustic parameters were analyzed with the Praat software (Boersma &Weenink, 2014) using a script for statistical voice analysis (Petersen, 2013). For visual inspection of the fundamental frequency (F0) distribution, histograms were generated for all subjects and all tasks. Average speaking pitch was analyzed through mean voice F0 in Hertz (Hz). Variation in F0 was measured both as the standard deviation (SD) of the average F0 and as the range in Hz, encompassing the middle 50% of F0 variation (middle 50% F0 range), thus reflecting intonation movements in terms of a quantitative measure where extreme values are omitted. Mean level difference in decibels (dB) between the first and second harmonics (H1-H2) of all voiced segments was measured for the here tasks. A high value of H1-H2 reflects a low grade of glottal compression and hyperfunctional voice quality. In all audio signals, laughter (5 incidences) and throat clearing (1 incidence) were edited out to ensure that the acoustic measures would reflect only speech. Mean F0 was identified with autocorrelation analysis in the range from 60 to 400 Hz. It was ensured that click sounds from the response buttons were not included in the autocorrelation analysis. The time step in the H1-H2 measurements was 0.05 seconds.

Statistical analysis

Examination of the data showed that response accuracy of the secondary task was near ceiling. The primary dependent variable of the study was the response times. Twenty-three of the 1080 trials had response times of 0 seconds. These were likely due to error responses and were removed from the dataset. Furthermore, 17 trials with response times lower than 200 milliseconds were excluded from further analysis, as this is considerably shorter than visual uptake and response. Data loss was thus lower than 4%. The remaining observations were skewed, so a log-transform was applied and these values were analyzed using a generalized linear mixed-effects regression model. This form of regression examines the extent to which a dependent variable, in this case log-transformed response times, is related via a link function (e.g., the logit function), to a combination of predictors. In our regression model we used a bottom-up method of analysis, where principal

components based on the aggregate questionnaire responses, trial order and speaking task were introduced in that order as fixed terms. Variables and interactions were omitted from the model if they were not significant at p>0.05. Stimulus and participant were included as crossed random effects. In addition, we included random slopes and levels corresponding to the fixed effects structure of the analysis. This allowed for the possibility that variables that were significant overall could also vary in their effect between participants.

In order to explore the structure of the questionnaire responses we conducted separate principal component analyses (PCA) on the responses from each task. Thereafter PCA was performed on the questionnaire responses from all three tasks in order to derive components that could be used as predictors in the mixed-effects model. This was because the original questionnaire responses were highly correlated and could not be used for this. Cronbach's alpha for the separate conditions was also used to assess the internal consistency and reliability of the responses. In addition, each of the acoustic parameters was analyzed with one-way repeated measures ANOVA with task as the independent variable. Analyses were performed with SPSS v 22 and R (R Development Core Team 2014), using the packages lme4 1.1 (Bates et al., 2014), psych 1.4 (Revelle, 2014) and lmerTEST 2.0 (Kuznetsova et al., 2015).

Results

Secondary task responses

The mean proportion of accuracy of the secondary task responses was found to be between 0.94 and 0.97 across the three speech conditions (see Figure 2, left panel). Thus the visual discrimination task seemed to have met the criterion for a secondary task as it proved to be cognitively simple (Paas et al., 2003).

The mean response times showed a clear difference across speech conditions (see Figure 2, right panel), being fastest for the rote speech (A), somewhat longer for the semi-spontaneous speech (C) and longest for the set form sentence speech (B).

Figure 2 about here

Subjective ratings

In the self-rating questionnaire, a higher score reflected a higher degree of perceived cognitive load with 40 as the maximum total score for all items (5 was the highest score on each of the eight items). The mean total score for each task is shown in Figure 3 and this can be seen to mirror the mean response times as task A had the lowest score, task B the highest, and task C was in-between. The internal consistency of the questionnaire was also assessed with Cronbach's alpha. Cronbach's alpha was 0.77 for task A; 0.88 for task B; and 0.95 for task C. These can be compared to 1 which represents perfect reliability (Crocker & Algina, 1986). Corrected item-total correlations, which are the correlations between each individual item and the composite score from all other items, were high (>0.7) for all items except 7 and 8 from task A and B.

Figure 3 about here

Loadings from the PCA of the separate tasks are displayed in Table 1. For task A and B, two

principal components were found that together accounted for 64% and 74%, respectively, of the total variance. PCA of the responses to task C identified only one component, which accounted for 77% of the variance. The first six questionnaire items from both Task A and B contributed considerably to the first component. Item 8 had the highest factor loadings for the second component, indicating that this item, which addressed the distraction imposed by the secondary task measured a different construct, which was not evident in the responses to task C.

Table 1 about here

PCA of the aggregated questionnaire responses (across all three tasks) revealed that 3 components accounted for 71% of the total variance. The factor loadings from this analysis are shown in Appendix 2 and reveal that item responses from task C contributed to the first factor, item responses from task A contributed to the second factor and item responses from task B contributed to the third factor. The principal components from this analysis were subsequently used in the mixed-effects model to examine associative links between the questionnaire responses and the response time data. It is advantageous to use principal components as predictors in regression models when the original variables, in this case the questionnaire responses, are highly collinear. In such cases the use of PCA reduces collinearity, which can be detrimental to the regression model.

Mixed-effects regression model

The log-transformed response times to all secondary task stimuli were used as the dependent variable in a mixed-effects model. The model is summarized in Table 2 and shows significant effects of speech task and trial order and an interaction between these variables. The direction of this interaction shows that response times for task B increased as the experiment progressed with an effect size of 647 ms, while for the other two tasks it decreased with effect sizes of 495 ms for task A and 485 ms for task C (see Figure 4). In addition to the substantial effect sizes for trial, the figure illustrates considerable differences between the different tasks, particularly later in the experiment.

As we were also interested in the association between the participants' subjective evaluation of their performance and their response times, we included the principal components based on the questionnaire from all blocks. The second principal component (PC) was not significant (F=0.6, p=0.58) and was therefore excluded from a candidate model. The non-linear effects of the first and third PCs indicate that there was an association between the response times and the subjective ratings, but interpreting the nature of these associations is problematic because the original questionnaire responses showed no clear pattern in their loadings on the PCs. Because of this, the PCs serve mostly to control this aspect and to establish the association between performance on the task and subjective ratings. With effect sizes of 824 ms for PC1 and 429 ms for PC3, it is clear that the subjective ratings are indirect but important predictors of response time.

Table 2 about here

Figure 4 about here

Acoustic characteristics

The fundamental frequency distributions for the three speech tasks can be seen in the histograms in Figures 5a and 5b. Almost all subjects had a bimodal distribution with the higher peak appearing at approximately 200 Hz and the lower peak an octave below. This pattern is commonly seen in the speech of young Danish women, where the peak at around 100 Hz corresponds to intermittent vocal fry and the Danish phenomenon of 'stød', which is a glottal articulatory gesture of vocal fold adduction or creaky voice (Fischer-Jørgensen,1989 a, b). The highest peak represents the most frequently occurring fundamental frequency in the sequence. A more narrow visual pattern of the F0 distributions can be seen for task A (reciting days of the week), representing a more monotonous intonation pattern compared to task B and C. This was confirmed by the measurement of middle 50% F0 range, where the lowest values were seen in task A, somewhat higher in task C and the highest in task B. The results of the acoustical analysis are summarized in Table 3.

Mean F0 was not effected by speech task ($F_{(2,22)}= .210$, $p = .812$) according to the ANOVA. However, SD of F0 and middle 50% F0 range, both measures reflecting variation in pitch, showed a significant effect of speech task (SD of F0: $F_{(2.22)}=19.069$, $p< .001$), and middle 50% F0 range: $F_{(2.22)}=7.802$, $p= .003$). Also, H1-H2 showed a significant effect of task ($F_{(2,22)}=12.843$, $p<.001$), with highest values in speech condition A (i.e. lowest degree of glottal compression), somewhat lower in C and lowest in B.

Figure 5a about here

Figure 5b about here

Table 3 about here

Discussion

The present investigation was an attempt to document with empirical data a hierarchical structure among three speech activities with regard to their level of cognitive load. This is a factor assumed to be of pedagogical relevance to the order in which speech and language pathologists select exercises, when a new speech behavior is ultimately to be automatized in a patient's spontaneous speech. Although neurological mechanisms behind various genres of speech are not yet completely understood, studies of lesioned patients indicate that vocal behavior is hierarchically organized and that different speech activities engage cerebral networks in unique combinations. For example, speech that is repeated and speech that is read aloud seems to require the least amount of attention or effort and has been shown to be associated with remarkable voice and speech improvements as compared to spontaneous speech, for example in Parkinsonian dysarthria (Kempler & Van Lancker, 2002), Huntington's disease (Vogel et al. 2014), spasmodic dysphonia (Roy, Gouse, Mauszycki, Merrill & Smith, 2005) and stuttering (Ludlow & Loucks, 2003). Rote speech typically refers to over-learned, recited or formulaic utterances of various kinds, including counting, familiar lyrics

and nursery rhymes. Rote speech is functionally represented in both right and left hemispheres and is thought to be linked to subcortical processing (Kreiman & Sidtis, 2011).

In the present investigation, both reaction times and subjective ratings were significantly lower for (A) rote speech than for (C) semi-spontaneous speech. This low level of cognitive load found in rote speech confirms that it should be well suited to an initial phase of therapy, thus allowing 'double focus' where the patient can attend to the sensory-perceptual feedback from speech production, which is important in ultimate habit formation and habit change in real-life speech contexts. While patients mechanically recite the weekdays, the speech and language pathologist can verbally guide them to interpret their intrinsic feedback and focus on functional production aspects of the voice, e.g. relaxation in the neck musculature. Such attention, analysis and cognitive activation of the patient may play a role in the inhibition of old habits and are believed to facilitate new motor and behavioral learning in terms of new schema construction and ultimately, automization in real-life situations (Iwarsson, 2014; Ohlsson, 2014). Spontaneous speech on the other hand, also called 'novel speech' or 'voluntary speech', refers to newly created, extemporaneous utterances. Such speech can be assumed to involve a larger amount of communicative effort and attention to content than rote speech. The term 'semi-spontaneous speech' was used in the present study to describe a condition in which the subjects were required to speak on a specific subject, rather than truly spontaneously initiating speech. The higher degree of cognitive effort found in this condition confirms clinical experience, that describing and narrating tasks can easily distract the patient from maintaining dual focus of attention on both content and revised behavior during production. In the same way as people who stutter are more vulnerable to increased cognitive load because cognitive load has a disrupting effect on speech language planning (Jones, Fox & Jacewicz, 2012), patients attending behavioral voice therapy may have a better chance of successfully adopting the new behavior in exercises with a low degree of distraction due to content.

Both reaction times and subjective ratings were highest for the sentences in set form (task B), i.e., to follow a set sentence structure and to add an element of free choice. This result was unexpected, as we anticipated that this task would incur only a marginal increase in cognitive load as compared to reciting weekdays alone. Also, clinical experience has indicated that this repetitive pattern of a short sentence (about 6 syllables) seems to be a safe task for the gentle introduction of 'double focus', with only a single element of free choice (Iwarsson, 2014). The mixed-effects model showed that this task differed significantly from the other two tasks, in general and with regard to the effect of trial on response times. Instead of response times decreasing with trial number, as they did in task A and C, task B showed an increase (see Figure 4). This may have reflected that subjects successively ran out of new ideas for the free choice activity-element of these sentences. This was confirmed during the post hoc analysis of the sound files as there were audible hesitations before this specific element. Moreover, it is possible that incidences of vocal fry in the filled pauses of these hesitations contributed to the low values of mean H1-H2 for this task.

Many studies have used measurements of the speech output itself as dependent variables affected by variations in task difficulty, time pressure, emotional state and cognitive load (Müller et al. 2001; Lay & Paivio, 1969; Bořil, Kleinschmidt, Boyraz, & Hansen, 1996; Huttunen, Keränen, Väyrynen, Pääkkönen, Leino, 2011). For example, increased mean F0, articulation rate, number of disfluencies as well as decreased duration of pauses and number of syllables per utterance have been reported due to increased time pressure and cognitive load. Mitchell and Hoit (1996) studied speech breathing and fluency-related behavior in two speech tasks differing in cognitive-linguistic processing. They showed that a higher demand condition of extemporaneous speech without a written outline, was associated with shorter phrases and a slower speaking rate when compared to pre-planned discourse with a written outline, referred to as a lower demand condition. They also found a lower average lung volume expended per syllable in extemporaneous speech, which is in accordance with our findings of a reduced H1-H2 difference in this task, as both of these measures reflect a greater level of glottal compression.

In the present study, we found significant effects of task for two of the acoustic measures; intonation variation and H1-H2. However, it should be noted that differences in speech behavior are not necessarily induced by cognitive load, but could also be affected by the inherent structure of the speaking task. For example, the recitation of weekdays performed in task A is a rather monotonous task where speakers could be expected to restrict variations in pitch, an observation which was borne out in the low middle 50% F0 range values recorded during this task (see Table 3). Such a monotonous task can be assumed to help the patient in an initial phase of voice therapy, for instance, by sustaining focus on the maintenance of a substantial transglottal airflow throughout phonation.

Subjective ratings of perceived task difficulty from a questionnaire administered directly after each condition reflected the response times of the secondary visual task. Mean response times for task A were 1140 ms and the mean questionnaire total was 13; for task B mean response times were 2203 ms and the mean questionnaire total was 30.5; and, for task C mean response times were 1799 ms and the mean questionnaire total was 21.5. The association between subjective ratings of task difficulty and response times on the secondary task was further supported as PCs derived from the questionnaire were found to be significant predictors of the response time data, indicating that subjects could accurately assess and report the cognitive load involved in the performance of tasks.

While PCA with so few items and responses should be interpreted with caution, analysis of the questionnaire data from the individual tasks showed that two factors accounted for the majority of the variance in tasks A and B. The factor loadings of these suggest that there were two underlying constructs, namely i) the cognitive demands required by the speech task, and ii) the distraction caused by the secondary task. It is interesting to note that the task C questionnaire responses yielded only one component. This may reflect that the visual secondary task was perceived to be less distracting during the production of semi-spontaneous speech than during the production of rote speech and on-line sentences in set form.

Future perspectives regarding research on cognitive load include a need to investigate empirically, various loading factors involved in speech. A traditional definition of cognitive load includes the mental effort required by a task in terms of working memory load (Yap, 2012). However, for speech activities cognitive load may be better framed in terms of attention. It seems highly relevant to assume involvement by content attention, motor-sensory attention, perceptual factors, linguistic structure of the task, emotional content and communicative factors such as eye contact and turn taking. These factors and their contributions to the general task difficulty need to be investigated and described in closer detail.

Conclusion

The present investigation has shown that in a homogenous and healthy group of female subjects, three speech tasks commonly used as voice therapy carry-over exercises varied as indexed by response time on a secondary visual discrimination task. This suggests that the exercises differed in the innate degree of cognitive load that they demanded. Furthermore, this supports the therapy idea to successively introduce speech tasks with increasing cognitive challenge, in order to ultimately acquire automaticity in real-life situations including communicative distractors. The speech acoustic measures of intonation variation and glottal compression also differed significantly between the three speech tasks. In addition, mean subjective ratings of self-reported difficulty showed a pattern that mirrored the mean response time data from the three tasks. The mixed-effects regression model supported an associative link between the response times and the subjective ratings, indicating that clients are likely to be aware of the cognitive load imposed by a given speech task and that their subjective judgment may be useful in therapy settings.

References

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. Educational Psychology Review, 22(4):425-438.

Baddeley, A. D., & Salamé, P. (1986). The unattended speech effect: perception or memory? Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 525–529.

Bates, D., Maechler M., Bolker, B., Walker, S. Christensen, R. H. B., Singmann, H., & Dai, B. (2014). <http://cran.r-project.org/web/packages/lme4/lme4.pdf>

Boersma, P., & Weenink, D. Praat: doing phonetics by computer (Version 5.4.05) [computer program]. 2014. Available at: http://www.praat.org/. Accessed Jan 20, 2015.

Bořil, H., Kleinschmidt, T., Boyraz, P., & Hansen, J.H.L. (1996). Impact of cognitive load and frustration on drivers' speech. Journal of Acoustical Society of America, 127. (Online published 2010)

Chen, J., Epps, J., Ruiz, N., & Chen, F. Eye activity as a measure of human mental effort in HCI. (2011). In Proceedings of International Conference on Intelligent User Interfaces, 315-318.

Crocker, L.M., & Algina, J. (1986) *Introduction to classical and modern test theory*, Holt, Rinehart, and Winston, Florida. .

Fischer-Jørgensen, E. (1989a). A phonetic study of the stød in standard Danish. Finland: Phonetics, University of Turku.

Fischer-Jørgensen, E. (1989b). The stød in standard Danish. Fonetica, 46, 1–59.

Huttunen, K., Keränen, H., Väyrynen, E., Pääkkönen, R., & Leino, T. (2011). Effect of cognitive load on speech prosody in aviation: Evidence from military simulator flights. Applied Ergonomics, 42:348-357.

Ikehara, C., & Crosby, M. (2005). Assessing cognitive load with physiological sensors. In Proceedings of Hawaii International Conference on System Sciences, p. 295a.

Iwarsson, J. (2014). Facilitating behavioral learning and habit change in voice therapy—theoretic premises and practical strategies. Logopedics Phoniatrics Vocology, Early online, 1-8.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall Inc., Englewood Cliffs, New Jersey.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. Science, 154 (No. 3756):1538-1585.

Kempler, D., & Van Lancker, D. (2002). The effect of speech task on intelligibility in dysarthria: Case study of Parkinson's disease. Brain and Language, 80, 449–464.

Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception.* Wiley-Blackwell, Oxford, UK.

Kuznetsova, A. Brockhoff, P. B., & Christensen, R. H. B. (2015). http://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf

Lay, C. H., & Paivio, A. (1969). The effects of task difficulty and anxiety on hesitations in speech. Canadian Journal of Behavioural Science, 1(1): 25-37.

Ludlow, C.L., & Loucks, T. (2003). Stuttering: A dynamic motor control disorder. Journal of Fluency Disorders, 28, 273–295.

Luximon, A., & Goonetilleke, R.S. (2001). Simplified subjective workload assessment technique. Ergonomics, 44(3):229-243.

Mitchell, H. L., & Hoit, J. D. (1996). Cognitive-linguistic demands and speech breathing. Journal of Speech Hearing Research, 39(1):93-104.

Moss, S. (1997). Limitations in human cognition and attention: Elucidating and validating the concept of multiple resource pools. Thesis, Monash University Publications of the Higher Education Research Data Collection (HERDC), ID 1997090319.

Müller, C., Grossmann-Hutter, B., Jameson, A., Rummer, R., & Wittig, F. (2001). Recognizing time pressure and cognitive load on the basis of speech: An experimental study. Lecture Notes In Computer Science, Vol. 2109:24-33.

Ohlsson, A.C. (2014). Verbal Instruction Model (VIM) in voice therapy. LPV Forum, Logopedics Phoniatrics Vocology, Early Online: 1-6.

Paas, F., Tuovinen, E., Tabbers, H., & Van Gerven, P.W.M. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 38(1):63-71.

Paas, F., Van Merriënboer, J.J.G., & Adam, J. (1994). Measurement of cognitive load in instructional research. Perceptual and Motor Skills, 79(1):419-430.

Petersen, N. R. (2013). Voice Analysis Tools [Praat script].

Posner, M.I. (1982). Cumulative Development of Attentional Theory. American Psychologist, 37(2): 168-179.

R Development Core Team. (2014). R: A language and environment for statistical computing [computer software manual]. Vienna, Austria.. <http://www.R-project.org/>.

Revelle, W. (2014). http://cran.r-project.org/package=psych

Roy, N., Gouse, M., Mauszycki, S.C., Merrill, R.M., & Smith, M.E. (2005). Task specificity in adductor spasmodic dysphonia versus muscle tension dysphonia, Laryngoscope, 115, 311–316.

Rubio, S., Dáz, E., Martín, J., & Puente, J.M. (2004). Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology: An International Review, 53(1), 61-86.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In Proceedings of SIGCHI Conference on Human Factors in Computing Systems, 2651-2656.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2):257-285.

Taib, R., Shi, Y.D., Choi, E., & Chen, F. (2007). Using pen input features as indices of cognitive load. In Proceedings of International Conference on Multimodal Interfaces, 315-318.

Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., & Schmidt, H.G. (2004). Memory load and the cognitive pupillary response in aging. Psychophysiology, 41(2):167-174.

Verdolini, K., & Lee, T. (2004) Optimizing motor learning in speech interventions. In C. Sapienza & J. Casper (Eds.), *Vocal rehabilitation for medical speech language pathology*. (p. 403–46) Pro-Ed; Austin, Texas. .

Vogel, A.P., Fletcher, J., & Maruff, P. (2014). The impact of task automaticity on speech in noise. Speech Communication, 65:1-8.

Yap, T. F. (2012). Speech Production Under Cognitive Load: Effects and classification. Ph D thesis from The University of New South Wales, School of Electrical Engineering and Telecommunications, Sydney, Australia.

Figures

Figure 1. Example of stimuli from the visual discrimination used as the secondary task. The 3 Japanese signs (Kanji ideograms) were presented on a computer screen where 2 were the same and 1 was different.
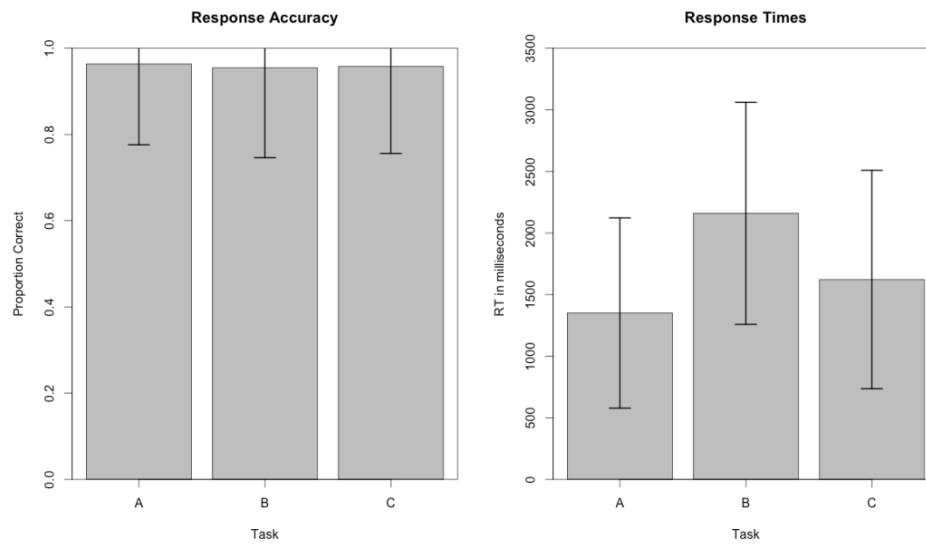
Figure 2. Results of the secondary task performance showing means (columns) and standard deviations (whiskers) for response accuracy (left panel) and response times (right panel) during the three speech tasks (A) rote speech, (B) sentences in a set form and (C)semi-spontaneous speech.
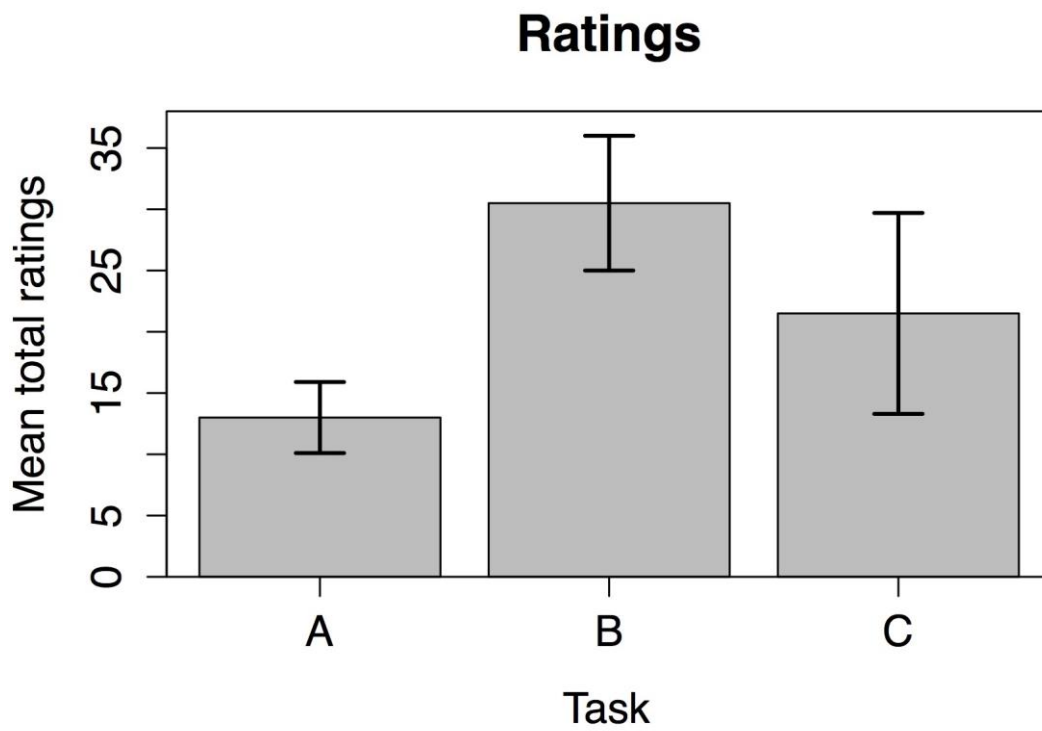
Figure 3. Mean total ratings across participants (bars) and S.D. (whiskers) for the questionnaire responses after (A) rote speech, (B) sentences in a set form and (C) semi-spontaneous speech.
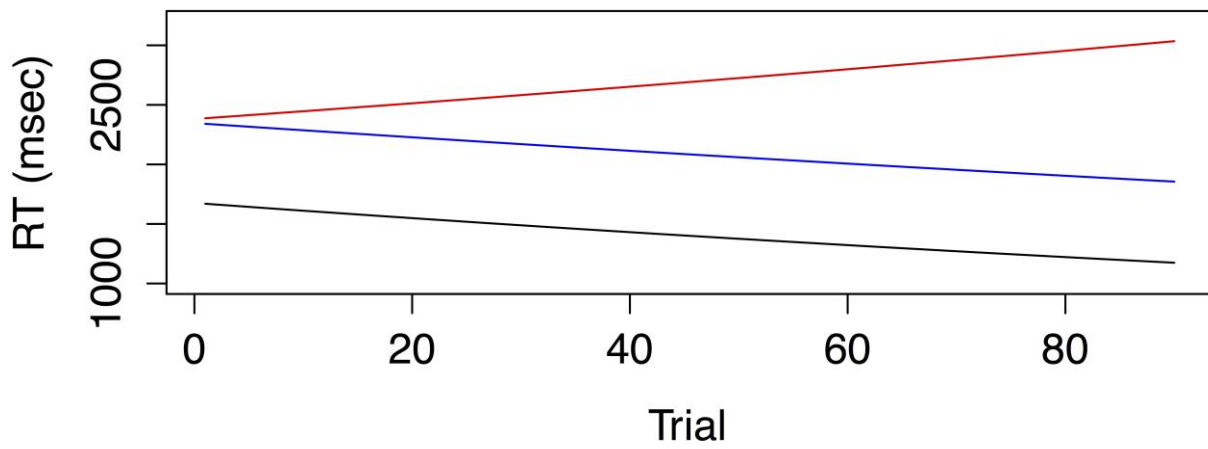
Figure 4. Effect of trial on the response times in the mixed-effects regression model according to task: task A rote speech (black), task B sentences in a set form (red) and task C semi-spontaneous speech (blue).The range of the horizontal axis (0-90) reflects that the task order was counterbalanced between participants.
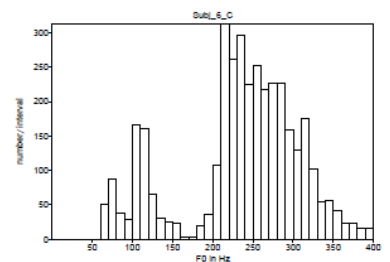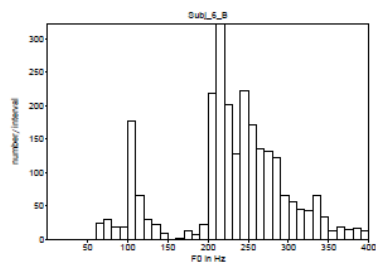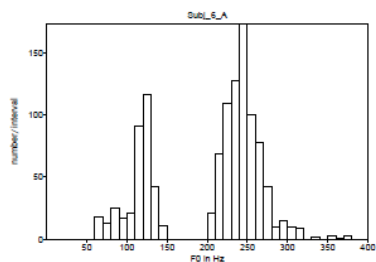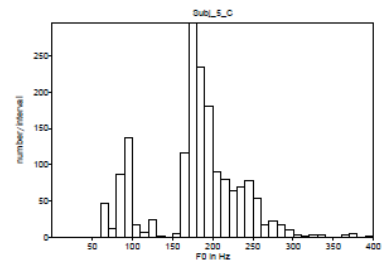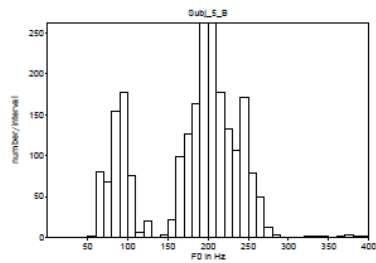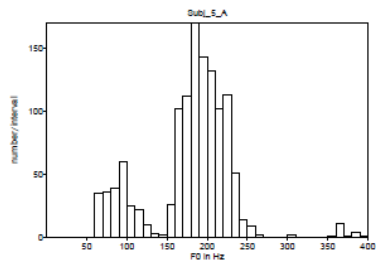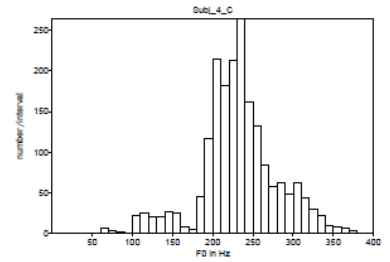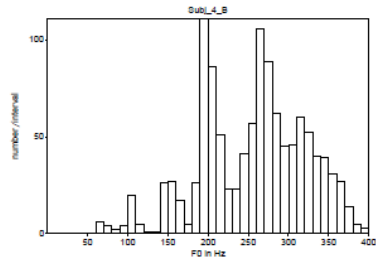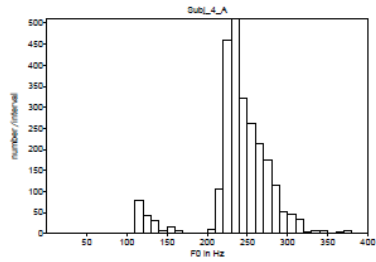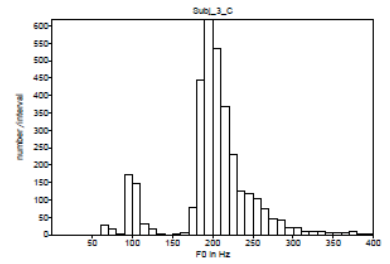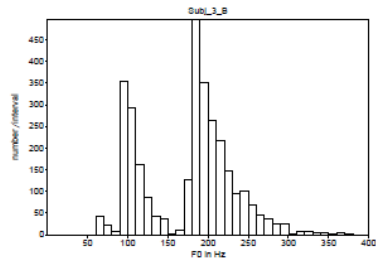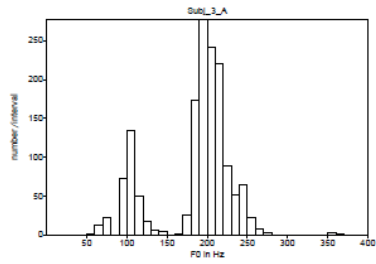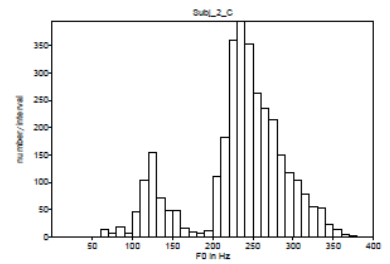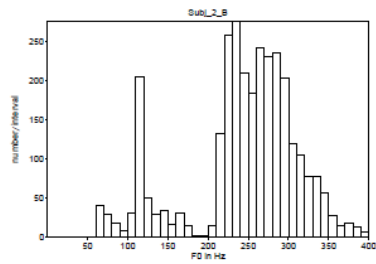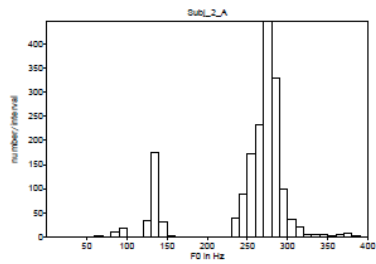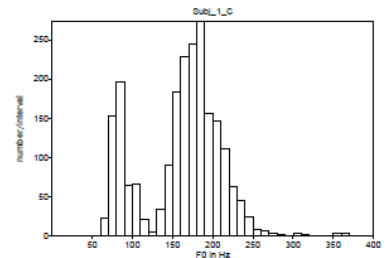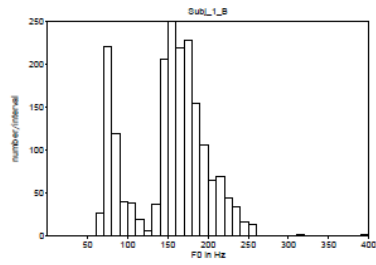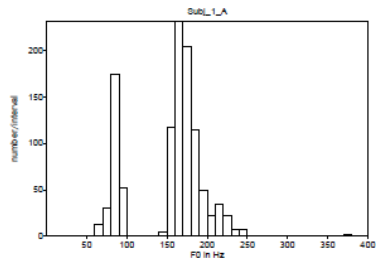
Figure 5a. Fundamental frequency distributions in Hz for subjects 1-6 (top to bottom) in the three speech tasks (left to right); task A rote speech, task B sentences in a set form and task C semi-spontaneous speech.
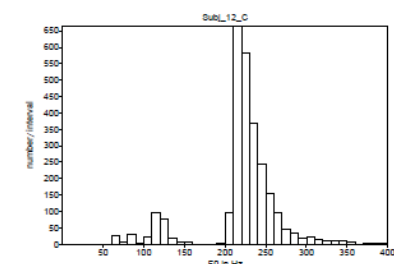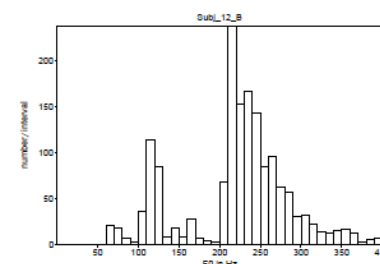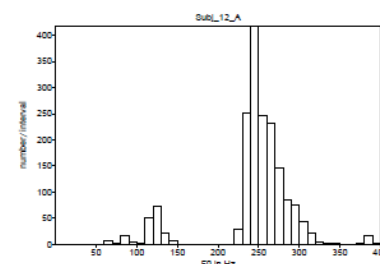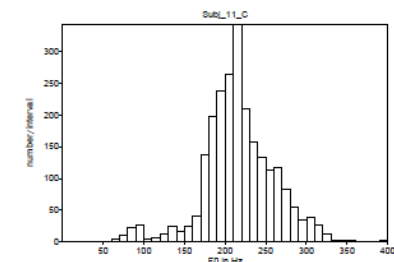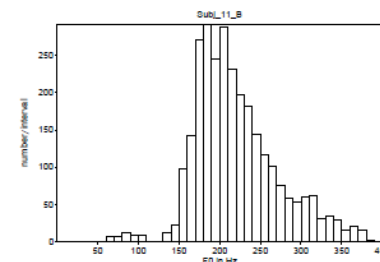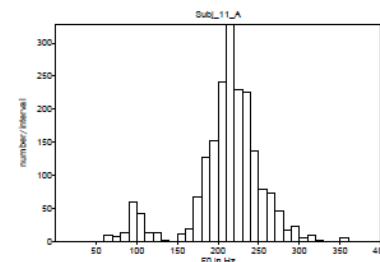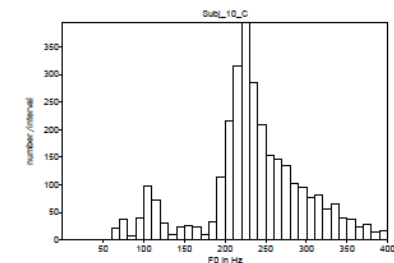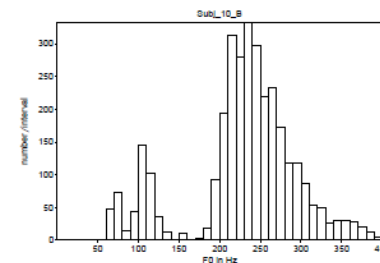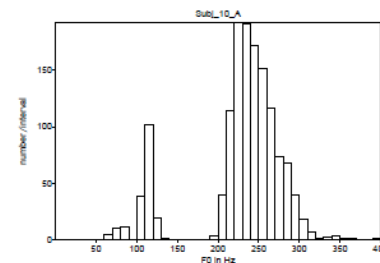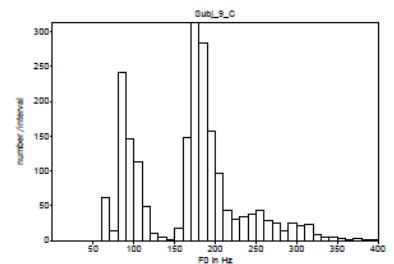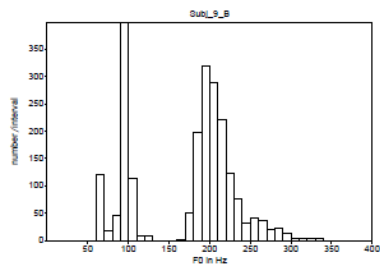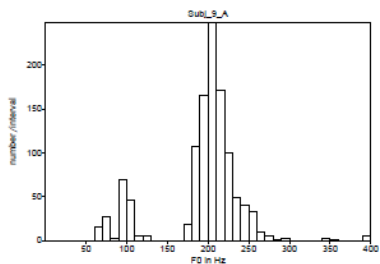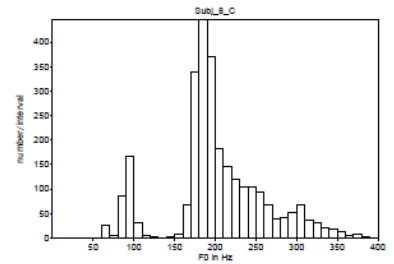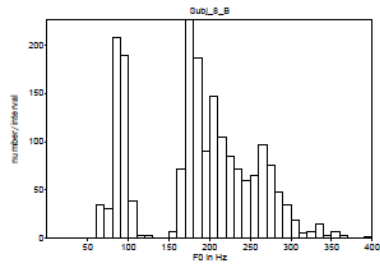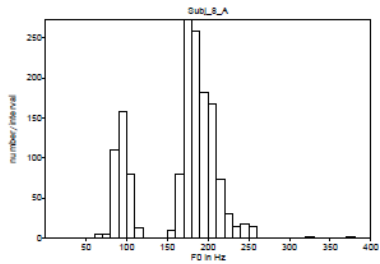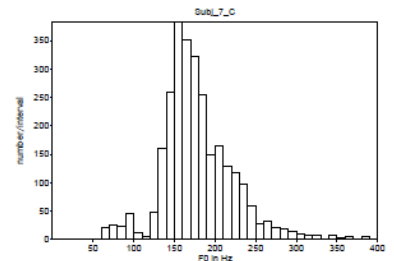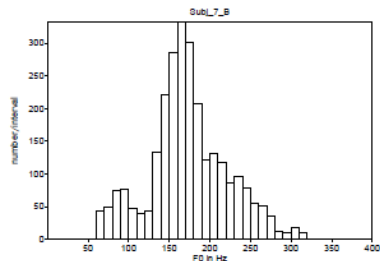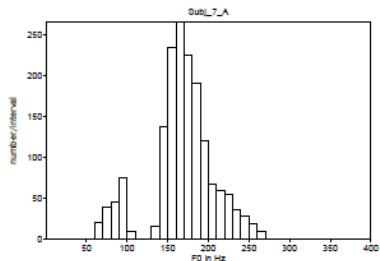
Figure 5b. Fundamental frequency distributions in Hz for subjects 7-12 (top to bottom) in the three speech tasks (left to right); task A rote speech, task B sentences in a set form and task C semi-spontaneous speech.

Tables

Table 1. Loadings of the original questions on the strongest principal components in the separate principal components analyses for each task. For task A and B, there are two principal components that account for most of the variance. For task C, one principal component is sufficient. Loadings lower than 0.6 are suppressed.

	Condition A		Condition B		Condition C
Item	PC1	PC2	PC1	PC2	PC
1	0.8	-	0.9	-	0.91
2	0.92	-	0.89	-	0.94
3	0.44	-	0.77	-	0.89
4	0.84	-	0.84	-	0.9
5	0.78	-	0.79	-	0.64
6	0.77	-	0.8	-	0.95
7	-	0.66	-	-	0.81
8	-	0.74	-	0.95	0.95
Variance explained	46%	18%	54%	20%	77%

Table 2. Summary of fixed effects in the mixed-effects model (n=1026 observations; REML criterion at convergence=632.1). Task A is the reference level for the factor task. Note that the linear terms are always included when the quadratic term is significant.

Predictors	Coefficient	Std. error	t	p
Intercept	7.32	0.06	113.12	<0.001
Task B	0.65	0.08	7.89	<0.001
Task C	0.39	0.04	8.68	<0.001
Trial order	-0.01	0.01	-1.84	<0.05
PC1	-0.06	0.01	-6.35	<0.001
PC1 non-linear	-0.01	0.01	-5.29	<0.001
PC3	-0.01	0.02	-0.6	n.s.
PC3 non-linear	-0.09	0.02	-4.65	<0.001
Task B x trial order	0.2	0.07	2.47	<0.05
Task C x trial order	0.03	0.06	0.63	n.s.

Table 3. Data from the acoustic analysis together with the main effect of task in a one way repeated measures ANOVA.

Acoustic variable	Task A	Task B	Task C	*df*	*F*	*p*
Mean F0 (Hz)	203	203	205	2, 22	.210	.812
SD of F0 (Hz)	49	61	55	2, 22	19,069	<.001
Middle 50% F0 range (Hz)	42	78	53	2, 22	7,802	.003
Mean H1-H2 (dB)	4	1	2	2, 22	12,843	<.001