Title:

What Does It Take to Stress a Word? Digital Manipulation of Stress Markers in Ataxic Dysarthria

Running Head:

Digital Manipulation of Stress Markers

Authors:

Anja Lowit[1], Tolulope Ijitona[2], Anja Kuschmann[1], Stephen Corson[3], John Soraghan[2]

[1]School of Psychological Sciences and Health, Strathclyde University

[2]Dept of Electrical and Electronic Engineering, Strathclyde University

[3]Department of Mathematics and Statistics, Strathclyde University

Corresponding Author:

Prof Anja Lowit

School of Psychological Sciences and Health

Strathclyde University

40 George St

Glasgow G1 1QE

Tel +44 (0)141 5483058

Email: a.lowit@strath.ac.uk

Abstract:

Background: Stress production is important for effective communication, but this skill is frequently impaired in people with motor speech disorders. The literature reports successful treatment of these deficits in this population, thus highlighting the therapeutic potential of this area. However, no specific guidance is currently available to clinicians on whether any of the stress markers are more effective than others, to what degree they have to be manipulated, and whether strategies need to differ according to the underlying symptoms. Aims: In order to provide detailed information on how stress production problems can be addressed, our study investigated (1) the minimum amount of change in a single stress marker necessary to achieve significant improvement in stress target identification; and (2) whether stress can be signalled more effectively with a combination of stress markers. Methods & Procedures: Data were sourced from a sentence stress task performed by 10 speakers with ataxic dysarthria and 10 healthy matched control participants. Fifteen utterances perceived as having incorrect stress patterns (no stress, all words stressed or inappropriate word stressed) were selected and digitally manipulated in a stepwise fashion based on typical speaker performance. Manipulations were performed on F0, intensity and duration, either in isolation or in combination with each other. In addition, pitch contours were modified for some utterances. 50 naïve listeners scored which word they perceived as being stressed. Outcomes & Results: Results showed that increases in duration and intensity at levels smaller than produced by our control participants resulted in significant improvements in listener accuracy. The effectiveness of F0 increases depended on the underlying error pattern. Overall intensity showed the most stable effects. Modifications of the pitch contour also resulted in significant improvements, but not to the same degree as amplification. Integration of two or more stress markers did not result in better results than manipulation of individual stress markers, unless they were combined with pitch contour modifications. Conclusions & Implications: Our results highlight the potential for improvement of stress production in speakers with motor speech disorders. The fact that individual parameter manipulation is as effective as combining them will facilitate the therapeutic process considerably, as will the result that amplification at lower levels than seen in

typical speakers is sufficient. The difference in results across utterance sets highlights the need to investigate the underlying error pattern in order to select the most effective compensatory strategy for clients.

What this paper adds:

Stress production is important for effective communication, but this skill is frequently impaired in clients with motor speech disorders. Clinicians still lack specific guidance how to address these issues.

Our paper provides structured data on the effects of acoustic stress marker manipulation on listener perception of which word is highlighted in an utterance.

The results show that a wide range of compensatory strategies are available to speakers to improve their stress production, even if they cannot modulate stress markers to the same degree as typical speakers. Clinicians need to consider the underlying error patterns carefully to identify the most appropriate compensatory strategy for an individual.

Introduction

There are two types of stress functions in spoken English, lexical stress and sentence stress. Sentence stress is a pragmatic-linguistic function used to emphasise particular words in an utterance. In doing so, attention is drawn to the most important part of an utterance, which supports effective communication between conversational partners by structuring discourse (e.g. Fine *et al.* 1991). In English, the position of stress is variable and speaker intent plays a significant role in stress placement, rendering the ability to mark stress appropriately an important one. This is reflected in observations that inappropriate use of sentence stress can reduce listener comprehension (e.g. Dooling 1974, Terken *et al.* 1987). Impairment of lexical stress can have similar impact on the comprehensibility of the speech signal. Lexical stress occurs within words and determines which syllable receives the highest prominence. This is important for a small number of minimal pair words in English whose meaning is exclusively distinguished by stress position (e.g. '**ob**ject vs ob'**ject**). Across an utterance, lexical stress furthermore contributes to the rhythmic pattern of the language and aids listeners identify word boundaries (Mattys *et al.* 1997), group lexical items (Dilley *et al.* 2008) and process syntactic and semantic information (Rothermich *et al.* 2012). Disturbances of stress placement, be this at the word or sentence level, are thus likely to impact on the listener's ability to understand the speech signal, particularly in instances where this is already reduced by other problems such as articulatory imprecisions as found in speakers with motor speech disorders. This highlights the importance of focusing on this area in the remediation of people with speech impairment.

Treatment of stress production deficits requires detailed knowledge of how this feature is produced. Research on typical speech indicates that acoustically, stress is primarily associated with three parameters: fundamental frequency (F0), intensity and duration (e.g. Fry 1958), although changes in vowel quality, voice quality and vocal tract state have also been identified as contributors to greater perceptual prominence (Laver 1980). F0 is generally regarded the most central marker for sentence stress (Fry 1958, Rietveld *et al.* 1985), followed by duration and intensity (e.g. Fry 1958, Morton *et al.* 1965), although some studies have reported different results, such as a greater reliance

on intensity, e.g. (Kochanski *et al.* 2005)). Whilst these studies describe overall patterns of behaviour, the perceptual salience of individual stress markers in an utterance will show a certain degree of variability, particularly in relation to target position. For example, other prosodic phenomena such as phrase final lengthening, and physiological constraints leading to a natural declination of F0 and reduction in intensity at the end of utterances can result in parameters being manipulated to a different degree depending on whether the stress target is in initial or final utterance position (Cooper *et al.* 1985). Lieberman (1960) furthermore introduced the concept of cue trading. He demonstrated that in cases where F0 was not higher on the stressed than the unstressed word, then intensity would compensate and become the primary stress cue. This phenomenon is thought to be of particular relevance in disordered speech as it can explain how some speakers are able to signal stress despite significant impairment in particular stress markers, and highlights the potential for the introduction of compensatory techniques.

The literature on stress production in populations with motor speech disorders has focused both on characterising the problems these speakers experience with lexical or sentence stress production as well as investigating treatment effects. Both developmental and acquired disorders have been studied, including cerebral palsy (CP) (Patel *et al.* 2009), traumatic or other types of brain injury (McHenry 1998, Simmons 1983, Wang *et al.* 2005, Yorkston *et al.* 1981, Yorkston *et al.* 1984), Parkinson's Disease (Cheang *et al.* 2007, Darkins *et al.* 1988, Gaviria 2015, Tykalova *et al.* 2014), Multiple Sclerosis (Hartelius *et al.* 1997), different types of ataxia (Liss *et al.* 1994, Lowit *et al.* 2010, 2014), and foreign accent syndrome (Kuschmann & Lowit, 2012). In terms of dysarthria types, the two most widely studied groups are ataxic and hypokinetic dysarthria.

In relation to hypokinetic dysarthria most investigations report diminished contrasts between stress and unstressed counterparts, i.e. speakers were able to raise F0, intensity or duration on the target, but not to the same degree as typical participants (Cheang *et al.* 2007, Tykalova *et al.* 2014), or showed no difference at all (Darkins *et al.* 1988). This finding is in line with the symptoms of monopitch and monoloudness commonly reported in PD and can lead to a perceptual impression of no word in the

utterance receiving particular focus. On the other hand, the literature on ataxic dysarthria presents a picture of stress being produced on the wrong target, or every word in the utterance being highlighted, particularly due to a lack of de-accentuation of unstressed elements (Lowit *et al.* 2012, 2014). Where speakers are able to signal stress correctly, they often rely on a limited number of acoustic parameters (Liss *et al.* 1994, Lowit *et al.* 2010, 2014, Patel *et al.* 2009, Yorkston *et al.* 1984). Yorkston *et al.* (1984), for example, describe three speakers with ataxic dysarthria, one of whom relied largely on duration to signal stress, whereas the others made more use of intensity and F0. A further feature that emerges from the literature is that speakers with dysarthria often overscale increases in stress markers. Whilst this can aid target identification for the listener, it can also result in their speech sounding unnatural (Yorkston *et al.* 1984) or actually have detrimental effects on speech perception if it impacts on the overall prosodic structure of the utterance (Liss *et al.* 1994).

Based on this previous research, it is clear that stress production strategies have to be carefully tailored to individual speakers' needs, taking into account their individual impairment patterns as well as potential effects of the therapeutic technique on other aspects of speech production. However, this advice can be difficult to implement for clinicians. Whilst perceptual analysis can highlight the presence of a stress production problem, detailed acoustic analysis is often necessary to identify the precise issues with the various stress markers underlying this problem. Clinicians rarely have the time or necessary acoustic analysis skills to perform such investigations. Furthermore, whilst the existing literature provides ample examples of impaired acoustic production patterns, not all of these studies have correlated this information with perceptual analyses to identify which can still result in successful target identification and which result in listeners being unable to locate the stressed word in an utterance. Furthermore, the reports that speakers could still signal stress successfully by significantly overscaling their production could give the impression that clients need to be asked to exaggerate their performance when cue trading is implemented as a strategy, leading to decreases in the naturalness of speech. Information on the necessary magnitude of parameter scaling as well as the optimum patterns of integration, i.e. how does the manipulation of a single parameter compare to

the integration of two or three in impacting perceptual accuracy, is thus essential for effective target setting in a therapeutic context.

To address this latter point, we aimed to provide more detailed information on the effects of acoustic manipulation of duration, intensity and F0 on listener's perceptions of sentence stress as part of a development of an automated system designed to provide effective feedback for clients during home-practice. We posed two specific questions, (1) what is the minimum amount of change in a single stress marker necessary to achieve significant improvement in target identification; and (2) can stress be signalled more effectively with a combination of stress markers than manipulations of individual parameters?

Methodology

Study Design

In order to answer our research questions, we performed a number of perceptual experiments which built on each other's results. First, the study data was selected on the basis of the perceptual results of a previous study of stress production in a group of people with ataxic dysarthria (Lowit *et al.* 2010). These data were then manipulated and presented to a new set of listeners to assess the effects of the manipulation to answer research question 1. On the basis of these results, further manipulations were performed and used in a second perceptual experiment to answer research question 2.

Two main paradigms to assess the effects of changes to speech production on listeners can be found in the motor speech disorder literature, those that ask speakers to change their performance (e.g. (Hammen *et al.* 1996, Tjaden *et al.* 2014), and those that manipulate the data digitally (Hammen *et al.* 1994, Laures *et al.* 1999, Watson *et al.* 2006). Whilst the former method will have greater face validity by being entirely based on naturally produced speech samples, the latter guarantees that the intended changes were actually performed and allows a higher degree of control over the extent of these changes. On this basis, the current study opted for the digital manipulation paradigm, as we aimed to

look at the effects of strictly controlled manipulations of individual parameters that would have been too difficult to control separately by the speakers to produce reliable answers to our questions.

Step1: Baseline Data Selection

The data used for this study constitute a subset of utterances recorded in a previous investigation for stress production in ten speakers with ataxic dysarthria and ten healthy gender, age and accent matched control speaker participants. All participants with ataxic dysarthria had a degenerative type of ataxia (undefined cerebellar, spino-cerebellar of various types or Friedreich's Ataxia) as diagnosed by the consulting neurologist. Presence of ataxic dysarthria was confirmed by the first author and severity of the speech disorder was expressed by the participants' intelligibility scores (see Table 1). These had been derived from visual analogue scale judgements of a reading passage (Cinderella Passage, The IViE Corpus) by seven trained listeners. All participants had sufficient cognitive, visual and hearing ability to complete the study tasks, and all were native monolingual speakers of English.

Insert TABLE 1 around here

As part of the test battery for the previous study, participants had completed a sentence stress task. Ten sentences with a Subject-Verb-Object-Adverbial (SVOA) structure, such as "The gardener grew roses in London", were produced in five different versions: (1) neutral intonation with all targets receiving similar degrees of stress, stress on the subject (2), object (3) or adverbial (4) (NB: the verb was never included as a target), and (5) contrastive stress on the subject. The different versions were elicited through a question-answer paradigm, such as "What happened" for (1), "Who grew roses in London?" for (2), or "Did the janitor grow roses in London?" for (5). This resulted in 50 sentences being produced per speaker, and 1000 in total for the study. All sentences contained a high number of voiced segments to allow for a reliable analysis of pitch contours. They were presented in a randomised order to avoid participants becoming used to a particular sentence structure.

The data were analysed acoustically to determine the speaker's strategies for stress production, by focusing on the three main stress markers duration, intensity and F0. In addition, a perceptual evaluation was performed by seven trained listeners who rated which word in the utterance had been stressed, or whether the whole utterance was produced neutrally with similar stress on all lexical items.

The perceptual analysis provided us with a list of utterances where listeners had struggled to identify the correct stress target. The acoustic analysis furthermore provided information about the magnitude of change produced by healthy control participants to mark stress. These were the basis for the current manipulation experiment outlined in steps 2 and 3.

Step 2: Data Manipulation and Perceptual Experiment 1

Listeners

50 naïve listeners evaluated the data of the current perceptual experiments. They were university students aged between 18 and 50 years who were unfamiliar with speech disorders. All were native speakers of English and did not have any hearing or speech impairment. To ensure that they understood the task they completed a practice experiment that included ten utterances with correctly produced stress patterns sourced from the healthy control samples. To progress, they had to correctly identify at least eight of these. Two volunteers had to be excluded from the experiment because their success rate in identifying the stress target was only 60% and 70% respectively, bringing the total number of recruits to 52. The majority of the remaining 50 listeners scored 100% in the practice experiment.

Speech materials – Experiment 1:

Based on the perceptual analysis results of the original study described above, we selected utterances that were misperceived by at least four of the seven listeners. This resulted in a sample of 124 utterances. These were checked for recording quality and grouped according to stress target position. Stimulus utterances were selected from this pool to arrive at a varied data set that covered all target positions across a range of speakers and sentences. The final utterance set contained five samples for each target position (i.e. 15 utterances in total), evenly sourced from across the stimulus sentences and speakers (with the exception of ATA 3 and 9 who had no problems marking stress).

Data Manipulation

Data manipulation procedures for this study were based on typical speaker behaviour and primarily focused on amplification of duration, intensity and F0 on the target words. However, the initial perceptual analysis revealed a number of different types of error patterns, i.e. the speaker was either unable to signal the target location (perceived either as all words stressed, or no stress), or they produced stress on an inappropriate word. For the "all stressed" pattern, the speakers had frequently produced an inappropriate pitch contour, such as placing an additional accent on a following or preceding word. We hypothesized that in these cases, the additional pitch accent might draw listeners' attention away from the target, even after this was amplified to a greater degree. We therefore introduced a further manipulation type to our procedure – pitch contour modification. This resulted in two sets of utterances, those that only underwent amplification (the AMP set) and those with inappropriate pitch contours (IPC set), which underwent amplification as well as modification.

The amplification procedure consisted of a step-wise increase in intervals of 25% for individual parameters. For F0 and intensity, this was aligned with healthy speaker performance, i.e. 25%, 50%, 75% and 100% of the mean control group difference between the target word and the following (for

T1) or preceding lexical item (for T2 and T3 positions). In addition, we also included a 150% F0 amplification condition for the IPC set to investigate whether higher than normal manipulations of this parameter might have similar effects as modifying the pitch contour on surrounding elements. Durational differences could not be measured within utterances as they were influenced by the phonetic structure of the individual target words which had not been controlled for during the development of the materials. For this parameter, we therefore based the increase on the actual duration of the word, i.e. a 100% increase relates to a doubling in length. The manipulations were conducted using a purpose built MATLAB algorithm that increased duration and intensity on the relevant items. For F0 amplification, we employed the speech synthesis function in Praat. The amplification was applied to the whole word, not just the stressed syllable, in order not to create any irregularities in duration, intensity or F0 within words.

For the modification procedure, we again used the Praat speech synthesis function to flatten the pitch accents on the inappropriately stressed word(s) to approximate the de-accentuation patterns typically produced by healthy participants. Figure 1 provides an example of F0 amplification at 75% as well as the modification procedure.

The above manipulations resulted in 211 stimuli for the first perceptual experiment (Table 2).

Insert TABLE 2 and FIGURE 1 around here

Data Presentation

The listening experiment was administered through Qualtrics. All test utterances were presented in visual form, both to reduce the effects of the speaker's intelligibility problems on stress perception and to allow for easy scoring of the perceived target word. After seeing the sentence on screen, listeners heard the corresponding sound file and were then asked to click on the word that they perceived to be stressed. Rest periods were introduced into the procedure every 30 minutes to avoid

loss of attention. To further counteract potential effects of performance decline over time, utterances were presented in random order so that no listener heard the same sequence of stimuli.

Step 3: Data manipulation and perceptual experiment 2:

In order to answer research question 2, a second perceptual experiment was conducted which investigated what impact the integration of two or more stress markers would have on listener success in identifying the target. In order to avoid confounding results of this experiment with single variable effects, the manipulations were mostly kept at the level below which amplification of the individual variables had resulted in significantly greater accuracy. The same parameter combinations were applied to the AMP and IPC utterances. In addition, we applied amplification of duration and intensity to the pitch modified versions in the IPC group.

The listening material for this experiment again included the original un-manipulated sound files, as well as the additional manipulations described above. By including the original files again, we were able to establish the level of natural response variation by the listeners. 229 audio samples were included in experiment 2 (Table 2), which took place around 4 weeks after experiment 1. The listeners, as well as data presentation and scoring procedures were identical to those of experiment 1.

Statistical Analysis:

The experiments yielded a single score for each data manipulation condition, i.e. the proportion of sentences for which the target word had been correctly identified. Two separate results were available for the baseline as the same utterances had been included in both experiments. The difference between these two values ranged between 2% and 9%, and there was no particular pattern regarding which target position or experiment resulted in better scores. The statistical analysis (Test of 2

Proportions, see below) revealed no significant difference between the values. For the purpose of statistical analysis, the mean of the two experiments was subsequently used as the baseline score.

To test for significance, we applied the Test of 2 Proportions, which is a hypothesis test to determine whether the difference between two proportions is significant. The two proportions represented the score for the un-manipulated sentences (baseline) and a particular degree and type of manipulation. For example, for durational manipulation, we could compare the baseline against 25, 50, 75 and 100% of duration increase and check what degree of manipulation resulted in a statistical significance. Due to the number of conditions included across the two experiments, this would have resulted in a vast number of statistical tests to be conducted, and thus a high likelihood of Type I errors. We therefore decided to only establish the minimum percentage increase necessary to reach significance at the 5% level, without calculating the exact p-values for the remaining results. This was sufficient as we were only interested in the minimum level of manipulation that resulted in a significant change rather than by how much each manipulation differed from the baseline. This method limited the number of necessary tests to an average of two per condition and thus did not require any statistical correction to be performed. Overall, results showed that depending on the condition, the necessary change required from baseline to reach significance fell between 11 and 14%. These figures are comparable to the statistically and/or clinically significant increases reported in other speech research (Deane *et al.* 2001, Van Nuffelen *et al.* 2010).

Results

Experiment 1:

The results for both experiments are summarised in Table 3. In addition, Figure 2 shows the results for the AMP and IPC utterance sets for duration, intensity and F0 for experiment 1.

As Figure 2 demonstrates, the baseline scores in the AMP utterances were slightly higher than was expected from our original perceptual analysis where at least four of the seven listeners had not been

able to identify the target successfully, whereas those for the IPC remain in line with the original findings. This suggests that there is more listener agreement when the wrong word has been stressed, than when a decision needs to be made whether all words or one particular one has been highlighted. Nevertheless, the results still show a general upward trend in listener accuracy in line with an increase in amplification across all three stress markers. There are some instances of worsening of scores (see e.g. AMP: D25 for T2 & T3, or I50 – I100 for T1), however, none of these decreases in accuracy were significantly different from the baseline score and could thus simply be a function of normal listener variance. Significant improvements in listener response were evident as a result of both amplification and modification procedures, with a number of patterns apparent relating to utterance type and target position.

Insert figure 2 and table 3 around here

Target Position Effects:

Baseline scores indicate that listeners had the greatest problems in relation to utterances with targets in medial position (T2), particularly in the IPC set. At the same time, these utterances benefited the most from parameter manipulation and showed similar, if not greater improvements in listener accuracy following amplification. The data furthermore demonstrate that the effectiveness of durational amplification was dependent on utterance position, i.e. across both the AMP and IPC set, T3 required the least amount of amplification, followed by T2 and then T1. This pattern was not as apparent for intensity or F0, which showed more even effects across the three utterance positions, e.g. T1, 2 & 3 results all became significantly higher at a 50% increase in intensity in the AMP set.

Utterance Type Effects:

Comparing the two types of utterances it is noticeable that less durational change was required to achieve significant improvements in the IPC set, whereas intensity results were largely comparable across the two. F0 manipulations show the greatest difference between the two sets, i.e. for AMP utterances, significant improvements were only achieved when F0 was increased to 100% of control mean, whereas changes as low as 25% could result in significant improvements in the IPC set (Table 3).

Experiment 2:

Experiment 2 revealed a similar pattern of results in that listener accuracy improved along with increased levels of manipulation, and the fundamental differences between target positions remained. In fact the positional effect was amplified in this experiment, as results showed no significant improvements for T1 at all in the AMP set, whereas most combinations of parameters showed significant results for T2 and T3. For IPC utterances T1 also showed improvements when F0 was manipulated alongside intensity or duration. However, the duration and intensity combination did not result in any significant changes, supporting the importance of F0 manipulation to achieve improvements in the IPC set identified in Experiment 1. The combination of contour modification plus amplification of duration or intensity was also highly successful and resulted in significant improvements throughout the various combinations except for the durational increase at 25%.

A comparison of the results of the two experiments reveals that the combination of parameters did not result in further overall improvement, i.e. the highest scores achieved with single parameter manipulation were not exceeded by adding a further parameter, or increasing all three. The only exception to this result is the contour modification condition, where results were considerably increased (to as far as 99% accuracy) when amplification of duration and intensity was added to the change in F0 direction.

Discussion

This research project aimed to investigate what kind of change in duration, intensity and F0 was necessary to achieve improved target word identification by naïve listeners. For this purpose, we looked into two separate conditions, (1) manipulation of individual stress markers, and (2) combinations of these.

The results of both listening experiments reflect previous research on stress production in that increases in duration, intensity or F0 of words, either individually or in combination with each other all resulted in at least some improvement in stress target identification by our group of 50 listeners. The results also reflected (Cooper *et al.* 1985) work on target positional effects on stress markers, i.e. their finding that durational increases tend to be higher if the target is in initial or medial than in final position was mirrored in our findings that a greater increase in duration was necessary to achieve significant improvements in listener accuracy in T1 and T2 conditions. In contrast, our data did not confirm F0 to be the most salient of the stress markers, followed by intensity and duration as is stipulated in some of the literature (Fry 1958, Rietveld & Gussenhoven 1985). Rather, a greater degree of change (100% of control speaker mean) was required for F0 to achieve significant improvements in the AMP utterances. Intensity, on the other hand, showed the most stable effects both across utterance sets as well as target positions, requiring only about 50% of healthy control increase to produce a significant result. This result is more in line with Kochanski *et al.*'s (2005) work, and also fits well with treatment approaches such as LSVT (Ramig *et al.* 2004) which focus singularly on establishing greater control over intensity production of speech and could possibly be integrated into stress production therapy as well. Although the necessary changes to duration appear to be lower than for F0 (starting at 25% for T3 in the AMP utterances), one has to consider that the durational increases in this experiment did not relate to a healthy level of performance change, but were absolute increases (i.e. up to a doubling of duration in the 100% condition). This is likely to be in excess of normal

performance differences, and duration could thus be categorised as the least effective parameter for single feature manipulation in the AMP set, particularly when the target was at the beginning of the utterance. The situation was very different in the IPC set though, where manipulations of all three parameters had comparable results. These utterances were unlike any reported previously on healthy speakers though and thus cannot be compared to the literature.

Although no significance testing could be carried out, it appeared that targets in the T3 position tended to be identified more easily than in the other conditions, i.e. they achieved the highest scores at baseline as well across the manipulations. This could be due to the fact that both intensity and F0 tend to decline towards the end of the utterance (Cooper *et al.* 1985), and thus even small increases in these parameters would be noticeable to listeners. With regard to duration, the results are surprising though, as utterance final lengthening means that durations are naturally increased in this position. One would thus assume that any further increases to mark stress would have to be considerable to override this natural phenomenon. Instead, our data show that T3 targets required the smallest durational increase of all positions to reach significance. The current results do not offer any clear explanation for this fact. Overall, the positive results for T3 utterances are encouraging given that this is the most common target location in natural discourse (Laver 1994). Focusing on this type of utterances during treatment could thus mean that therapy targets are more easily achievable, and that they translate into everyday communication more readily.

From a clinical perspective, it was reassuring to note that on the whole, as suggested by the literature (Fry 1958, Lieberman 1960), manipulations of either of the parameters showed positive effects. This confirms that any client with the ability to modify duration, intensity or F0 should have the potential to improve their ability to signal stress in an utterance. Furthermore, in many cases, significant improvements could be observed with lower levels of increase than was produced by the healthy control speakers. This could partly be due to the fact that our control speakers appeared to exaggerate their responses as part of the stress production paradigm, i.e. they increased their stress markers to a

higher degree than is necessary in naturally flowing discourse. An area that requires further investigation is whether manipulations beyond the control speaker mean will result in continued improvements. Our results suggest that this might be the case based on the results of the 150% F0 increase paradigm, however, no information is available on the other parameters or the AMP utterance set. Reports from the literature also suggest that this could be the case, e.g. (Yorkston *et al.* 1981) report that one of their speakers who exaggerated intensity increases was 100% successful in signalling stress target position. At the same time, they raise concerns about the bizarreness ratings of the speaker. This suggests that a balance will need to be struck between the degree of success in marking stress and the resulting perceptual impression of the overall speech signal. Our results suggest that even with small changes, success can reach the high 80 % to low 90 % range, which might be acceptable to clients as a therapy goal.

In relation to the IPC utterance set, we hypothesised that increasing the stress markers on the actual target might not have much effect, as the listener's attention would be drawn to the inappropriate word by the pitch contour. Against our expectation, amplification of individual stress markers, at even lower levels than in the AMP condition, showed successful outcomes. In fact, the overall effects were greater than for the AMP set, i.e. the largest improvement in listener accuracy observed in the AMP utterances was 20%, whereas this reached up to 47% in the IPC set. Whilst one has to bear in mind that the baseline scores were lower in the IPC than the AMP utterances, providing more scope for improvement, the results still demonstrate that even utterances with relatively serious prosodic disturbances can be improved with small changes to production. Furthermore, contrary to the AMP set, IPC utterances resulted in better outcomes with the amplification of any of the three stress markers, thus providing more possibilities for compensation or cue trading. Finally, modification of the contour by changing a rising to a falling tone to resemble the de-accentuation patterns observed in the typical speakers also resulted in increased listener accuracy. Although the improvements were not as great as for F0 amplification, changing the direction of the pitch contour might be a viable alternative for clients with reduced pitch range if they retain sufficient control over their pitch

direction, especially as falling tones are easier to produce than rising ones. Further investigation is required to assess the feasibility of such an approach though, as it is currently unknown whether the lack of de-accentuation that leads to the inappropriate pitch contours is inherent to the motor speech disorder and thus not easily treatable, or a result of inappropriate compensatory techniques that could potentially be modified during therapy.

The results of experiment 2 were again slightly unexpected, as we hypothesised that combining stress markers would result in greater listener accuracy than modifying single parameters, given that healthy speakers are reported to modulate all three in any given context. Instead, we found significant evidence of cue trading, i.e. the employment of only one stress marker was able to compensate for the absence of modifications in the remaining two cf. (Lieberman 1960). From a clinical perspective, this is a positive outcome, as it will be easier for clients to concentrate on modulating a single parameter than two or three. The only cases where integration of parameters resulted in better outcomes than those observed in experiment 1 were when intensity or durational increases were added to the contour modification in the IPC utterances, particularly where the target was in final position with results close to 100% correct.

Clinical Implications

Our investigation has revealed a number of issues that should be considered by clinicians when working with clients with stress production problems due to motor speech disorders:

1. The clinician needs to establish which of the stress parameters a client can still manipulate and to what degree, in order to decide the best strategy for compensatory behaviour.

2. Focusing on a single stress marker appears to have similar outcomes for stress recognition as manipulating several together. This should facilitate treatment considerably, especially if the

client retains better control over one than the others and effective compensatory strategies can thus be implemented.

3. It is important to identify the underlying pattern of inappropriate stress marking (is the target word not highlighted sufficiently, or the wrong word highlighted), as this can influence the choice of therapeutic option:

    a. If the underlying problem is largely that targets are insufficiently highlighted, our results suggest that an increase in intensity can have the most beneficial effects, as it was equally effective across all three target positions at around 50% of what healthy speakers would produce.

    b. If, on the other hand, too many words are stressed, our data suggest that a combination of pitch contour correction and amplification of duration or intensity could significantly improve outcomes. However, if the client lacks sufficient control to produce a contour change, then amplification of any individual stress marker can still result in significant increase in listener understanding, albeit at lower levels.

A number of caveats need to be considered in the interpretation of the above results. First, one has to keep in mind that the current data have been digitally manipulated, rather than asking speakers to alter their speech production. Whilst this means that we still have no information to what degree speakers with speech disorders are able to produce the required changes and what techniques are most effective to help them to do so, the current paper has established an important basis for such treatment studies in highlighting what stress markers clinicians could focus on to support their clients in this endeavour.

Also, the stimulus sentence selection was purely based on the fact that less than 50% of our initial listener set had been unable to identify the stress target accurately. Our baseline results indicate that the current, larger group of listeners was more successful in identifying the target word and it is thus

likely that at least some stress parameters had already been manipulated to some degree. This could have resulted in greater manipulation of a particular parameter than we assumed, or in fact the integration of several parameters when we assumed only one had been altered. To minimise these problems, we ensured that we sourced the utterances from a variety of speakers and sentence types, and any effects caused by the source material should thus have been subsumed within the greater data set. In addition, by choosing actual disordered samples rather than more controlled, artificial materials as our basis, the current data represent more closely the types of issues clients can present with in clinic and what strategies will be useful to address their problems.

Furthermore, some acoustic manipulations, in particular pitch modifications at 100%, resulted in the modified version of the word sounding unnatural at times, which possibly directed the listeners' attention to it. However, our clinically most relevant outcome was that significant improvements could be achieved with lower levels of manipulation which did not sound unnatural at all. In addition, some of the exaggerated productions that are frequently elicited from clients in clinic are also often perceived as unnatural, and listeners' reactions might thus be similar to our digitally manipulated and naturally produced utterances (cf. Yorkston *et al.* 1981).

Finally, the current data are only based on speakers with ataxic dysarthria, and there is a question to what degree the results can be generalised to other speakers with motor speech disorder. This fact is to some part mitigated by the fact that our speakers showed a range of error patterns which can also be found in other types of dysarthria, and the nature of our results, which promote a personalised approach based on the nature of the stress production problem of individual clients. For example, the finding that only small amplifications in stress markers can lead to marked improvement in stress perception are actually more relevant to the impairments of monopitch and monoloudness associated with people with hypokinetic dysarthria, than the excessive stress production frequently associated with ataxic dysarthria. A further encouraging finding for this client group was the fact that loudness was one of the most effective parameters to use in light of the evidence we have from therapeutic

studies such as the LSVT literature (Ramig *et al.* 2004), which demonstrate that a variety of disordered speaker populations are able to achieve greater control over this speech feature.

Conclusion

This was the first research to systematically manipulate stress markers in disordered speech samples to identify the type and scale of prosodic adjustment necessary to mark stress successfully. The skill of highlighting elements in a string of speech, plays a considerable role in directing listeners' attention to the important aspects of our message, and also supports speech perception. Working on stress placement is thus an important therapeutic goal to support clients with reduced communication effectiveness. This paper has demonstrated potential manipulations to disordered speech samples that help listeners identify the location of the stress target, providing guidance to clinicians working in this area. In addition, we were able to provide information on the levels of change required across different parameters, utterance positions and error patterns, which will serve as a basis for target setting for future developments of technological treatment aids.

References

CHEANG, H. S., AND PELL, M. D., 2007, An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. Journal of Neurolinguistics, 20, 221-241.

COOPER, W. E., EADY, S. J., AND MUELLER, P. R., 1985, Acoustical aspects of contrastive stress in question-answer contexts. Journal of the Acoustical Society of America, 77, 2142-2156.

DARKINS, A. W., FROMKIN, V. A., AND BENSON, F., 1988, A characterization of the prosodic loss in Parkinson's disease. Brain and Language, 34, 315-327.

Deane, K. H., Whurr, R., Playford, E. D., Ben-Shlomo, Y., and Clarke, C. E. (2001). Speech and language therapy for dysarthria in Parkinson's disease: a comparison of techniques (Publication no. 10.1002/14651858). (CD002814).

DILLEY, L. C., AND MCAULEY, J. D., 2008, Distal prosodic context affects word segmentation and lexical processing. Journal of Memory and Language, 59, 294-311.

DOOLING, D. J., 1974, Rhythm and syntax in sentence perception. Journal of Verbal Learning and Verbal Behavior, 13, 255-264.

FINE, J., BARTOLUCCI, G., GINSBERG, G., AND SZATMARI, P., 1991, The use of intonation to communicate in pervasive developmental disorders. Journal of Child Psychology and Psychiatry, 32, 771-782.

FRY, D., 1958, Experiments in the perception of stress. Language and Speech, 1, 126-152.

GAVIRIA, A. M., 2015, Acoustic Realization of Contrastive Stress in Individuals with Parkinson's Disease. Unpublished, Louisiana State University and Agricultural and Mechanical College

HAMMEN, V. L., AND YORKSTON, K. M., 1996, Speech and Pause Characteristics following Speech Rate Reduction in Hypokinetic Dysarthria. Journal of Communication Disorders, 29, 429-445.

HAMMEN, V. L., YORKSTON, K. M., AND MINIFIE, F. D., 1994, Effects of Temporal Alterations on Speech-Intelligibility in Parkinsonian Dysarthria. Journal of Speech and Hearing Research, 37, 244-253.

HARTELIUS, L., WISING, C., AND NORD, L., 1997, Speech modification in dysarthria associated with

multiple sclerosis@ An intervention based on vocal efficiency, contrastive stress, and verbal

repair strategies. . Journal of Medical Speech - Language Pathology, 5, 113-140.

The IViE Corpus: English Intonation in the British Isles. Retrieved from

http://www.phon.ox.ac.uk/files/apps/IViE/index.php, 6 April 2008.

KOCHANSKI, G., GRABE, E., COLEMAN, J., AND ROSNER, B., 2005, Loudness predicts prominence:

Fundamental frequency lends little. Journal of the Acoustical Society of America, 118, 1038-

1054.

LAURES, J. S., AND WEISMER, G., 1999, The effects of a flattened fundamental frequency on intelligibility

at the sentence level. Journal of Speech, Language, and Hearing Research, 42, 1148-1156.

LAVER, J., 1980, The phonetic description of voice quality (Cambridge: Cambridge University Press).

LAVER, J., 1994, Principles of Phonetics (Cambridge: Cambridge University Press).

LIEBERMAN, P., 1960, Some acoustic correlates of word stress in American English The Journal of the

Acoustical Society of America, 32, 451-454.

LISS, J. M., AND WEISMER, G., 1994, Selected acoustic characteristics of contrastive stress production in

control geriatric, apraxic, and ataxic dysarthric speakers. Clinical Linguistics & Phonetics, 8,

45-66.

LOWIT, A., AND KUSCHMANN, A., 2012, Characterizing intonation deficit in motor speech disorders: an

autosegmental-metrical analysis of spontaneous speech in hypokinetic dysarthria, ataxic

dysarthria, and foreign accent syndrome. Journal of Speech, Language, and Hearing

Research, 55, 1472-1484.

LOWIT, A., KUSCHMANN, A., AND KAVANAGH, K., 2014, Phonological markers of sentence stress in ataxic

dysarthria and their relationship to perceptual cues. Journal of Communication Disorders,

50, 8-18.

LOWIT, A., KUSCHMANN, A., MACLEOD, J. M., SCHAEFFLER, F., AND MENNEN, I., 2010, Sentence Stress in

Ataxic Dysarthria - A Perceptual and Acoustic Study. Journal of Medical Speech - Language

Pathology, 18, 77-82.

MATTYS, S. L., AND SAMUEL, A. G., 1997, How Lexical Stress Affects Speech Segmentation and

Interactivity: Evidence from the Migration Paradigm. Journal of Memory and Language, 36,

87-116.

MCHENRY, M., 1998, The ability to effect intended stress following traumatic brain injury. Brain

Injury, 12, 495-503.

MORTON, J., AND JASSEM, W., 1965, Acoustic correlates of stress. Language and Speech, 8, 159-181.

PATEL, R., AND CAMPELLONE, P., 2009, Acoustic and Perceptual Cues to Contrastive Stress in Dysarthria.

Journal of Speech, Language, and Hearing Research, 52, 206-222.

RAMIG, L. O., FOX, C., AND SAPIR, S., 2004, Parkinson's disease: speech and voice disorders and their

treatment with the Lee Silverman Voice Treatment. Seminars in Speech and Language, 25,

169-180.

RIETVELD, T. C. M., AND GUSSENHOVEN, C., 1985, On the relation between pitch excursion size and

prominence. Journal of Phonetics, 13, 299-308.

ROTHERMICH, K., SCHMIDT-KASSOW, M., AND KOTZ, S. A., 2012, Rhythm's gonna get you: Regular meter

facilitates semantic sentence processing. Neuropsychologia, 50, 232-244.

SIMMONS, N. N., 1983 Acoustic analysis of ataxic dysarthria: An approach to monitoring treatment. In

W. Berry (Ed.) Clinical Dysarthria (Austin, TX: Pro-Ed), pp. 283-294

TERKEN, J., AND NOOTEBOOM, S. G., 1987, Opposite effects of accentuation and deaccentuation on

verification latencies for given and new information. Language and Cognitive Processes, 2,

145-163.

TJADEN, K., SUSSMAN, J. E., AND WILDING, G. E., 2014, Impact of clear, loud, and slow speech on scaled

intelligibility and speech severity in parkinson's disease and multiple sclerosis. Journal of

Speech, Language, and Hearing Research, 57, 779-792.

TYKALOVA, T., RUSZ, J., CMEJLA, R., RUZICKOVA, H., AND RUZICKA, E., 2014, Acoustic Investigation of Stress Patterns in Parkinson's Disease. Journal of Voice, 28, 129.e121-129.e128.

VAN NUFFELEN, G., DE BODT, M., VANDERWEGEN, J., VAN DE HEYNING, P., AND WUYTS, F., 2010, Effect of rate control on speech production and intelligibility in dysarthria. Folia Phoniatrica et logopaedica, 62, 110-119.

WANG, Y. T., KENT, R. D., DUFFY, J. R., AND THOMAS, J. E., 2005, Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress. J Commun Disord, 38, 231-260.

WATSON, P. J., AND HUGHES, D., 2006, The Relationship of Vocal Loudness Manipulation to Prosodic F0 and Durational Variables in Healthy Adults. Journal of Speech, Language, and Hearing Research, 49, 636-644.

YORKSTON, K. M., AND BEUKELMAN, D. R., 1981, Ataxic dysarthria: treatment sequences based on intelligibility and prosodic considerations. J Speech Hear Disord, 46, 398-404.

YORKSTON, K. M., BEUKELMAN, D. R., MINIFIE, F. D., AND SAPIR, S., 1984 Assessment of stress patterning. In M. R. MacNeil, J. C. Rosenbek, & A. E. Aronson (eds), The dysarthrias: Physiology, acoustics, perception, management (San Diego: College Hill Press), pp. 131-162

Table 1: Participant details for speakers with ataxic dysarthria (F=female, M=male; CA: cerebellar ataxia of undefined type, SCA: spinocerebellar ataxia & type, FA: Friedreich's Ataxia)

| speaker	age	gender	etiology	% intelligibility score
ATA1	46	M	CA	26
ATA2	60	F	CA	33
ATA3	28	M	FA	94
ATA4	52	F	CA	75
ATA5	28	F	FA	91
ATA6	65	F	SCA6	42
ATA7	72	M	CA	81
ATA8	51	M	CA	56
ATA9	56	M	SCA8	18
ATA10	57	F	FA	20
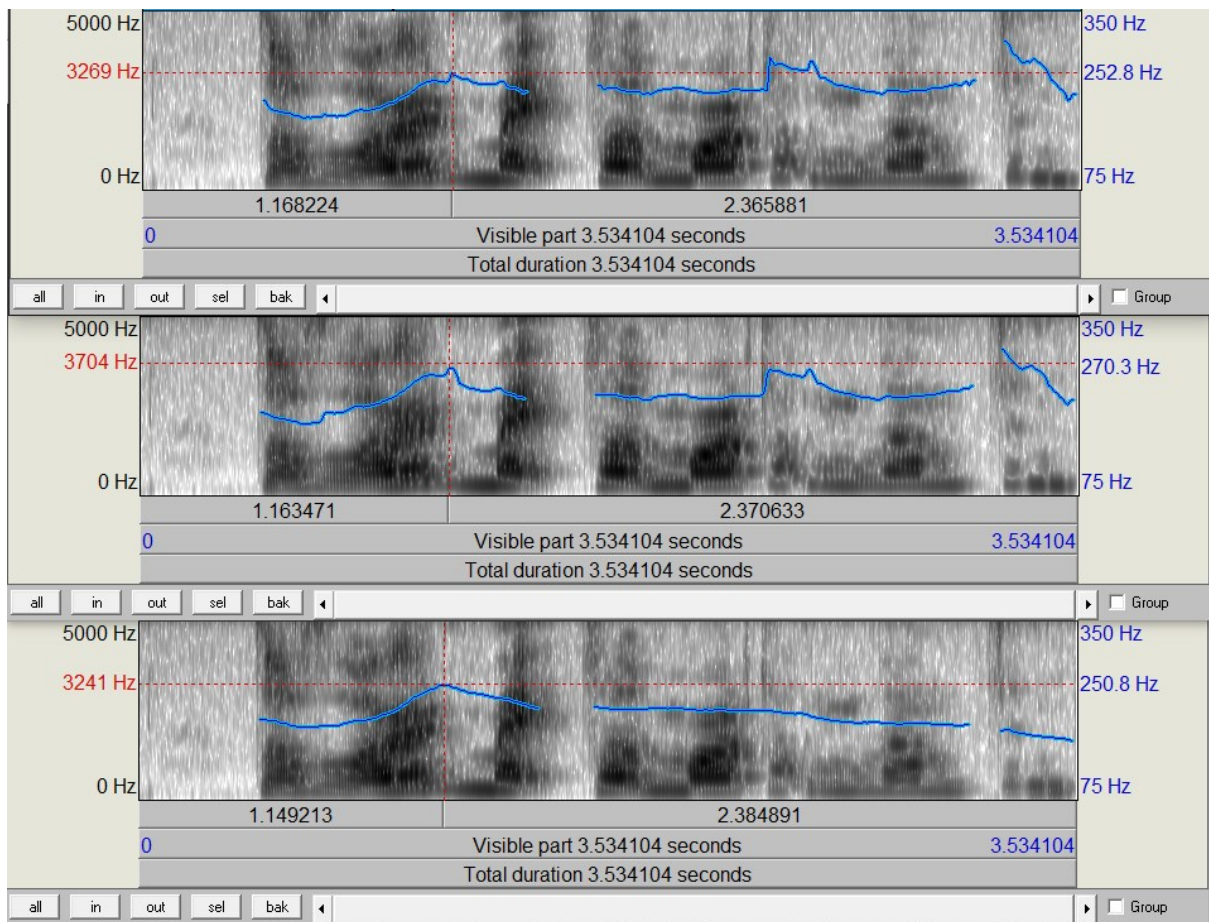
Figure 1: Example of F0 amplification and contour modification for "The **lawyer** met the model in London"; the unmodified version (top) shows peaks on "lawyer, model and London", in the amplified version (middle), the peak on "lawyer" was increased, and in the modified version (bottom) the inappropriate peaks on "model" and "London" were removed and the contour smoothed to resemble the typical de-accentuation pattern.

Table 2. Modification Conditions for the Two Perceptual Experiments

	Description of Modification	No of Stimuli	
		AMP	*IPC*
	Experiment 1		
1.	No modification	7	8
2.	Increase in Intensity (25%, 50%, 75%, 100%)	28	32
3.	Increase in F0 (25%, 50%, 75%, 100%)	28	32
4.	Increase in Duration (25%, 50%, 75%, 100%)	28	32
5.	Pitch Contour Modification	0	8
6.	Increase in F0 (150%)	0	8
	TOTAL	211	
	Experiment 2		
1.	No modification	7	8
2.	25% Duration & 25% Intensity	7	8
3.	25% Duration & 50% Intensity	7	8
4.	50% Duration & 25% Intensity	7	8
5.	75% F0 & 25 % Intensity	7	8
6.	75% F0 & 50% Intensity	7	8
7.	100% F0 & 25% Intensity	7	8
8.	25% Duration & 75% F0	7	8
9.	25% Duration & 100% F0	7	8
10.	50% Duration & 75% F0	7	8
11.	25% Duration, 25% Intensity & 75% F0	7	8
12.	Pitch Contour Modification and Intensity Increase (25%, 50%, 75% & 100%)	0	32
13.	Pitch Contour Modification and Duration Increase (25%, 50%, 75% & 100%)	0	32
	TOTAL	229	
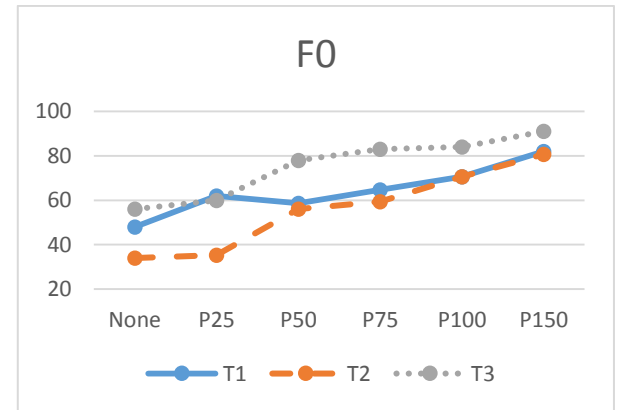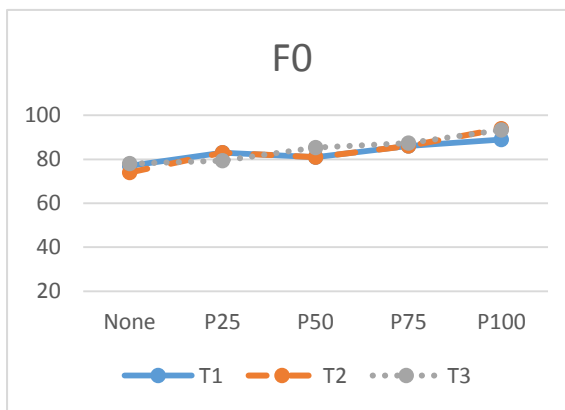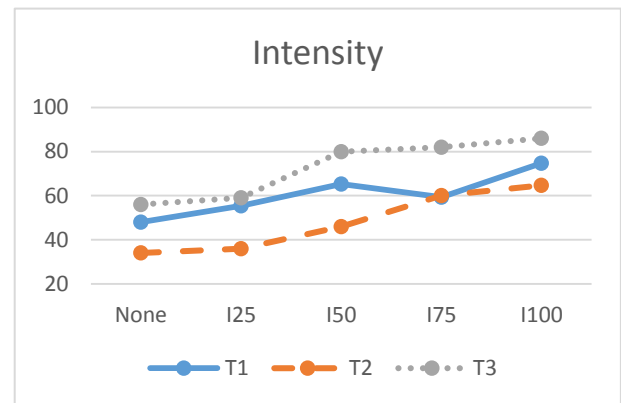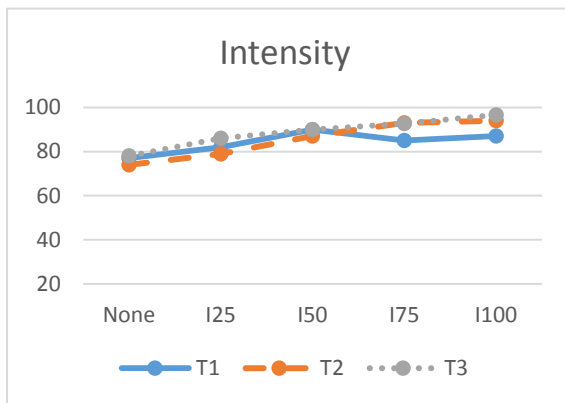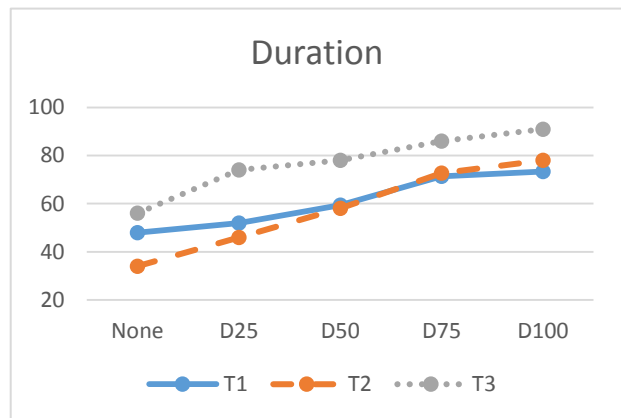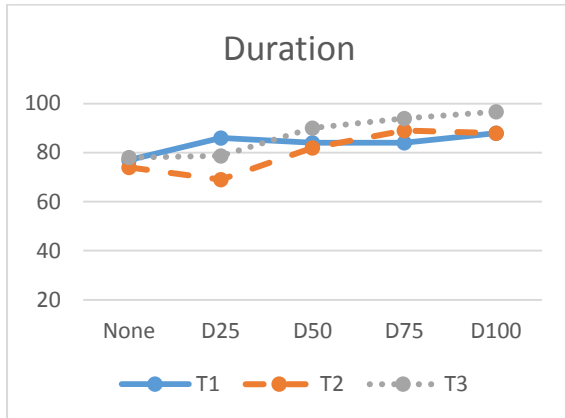
Figure 2: Results for duration, intensity and F0 amplifications for AMP and IPC utterances in Experiment 1. The x-axis shows the various amplification conditions, and the y-axis the percentage of utterances perceived with the correct stress pattern. Each graph shows results for initial (T1), medial (T2) and final stress targets (T3).

Experiment 1

AMP

	Baseline	D25	D50	D75	D100	I25	I50	I75	I100	P25	P50	P75	P100	P150	MOD
T1	77	86	84	84	88 .039	82	90 .012	85	87	83	81	86	89 .022		
T2	74	69	82	89	88 .010	79	87 .031	93	94	83	81	86 .032	94		
T3	78	79	90 .019	94	97	86	90 .019	93	97	79	85	87	93 .002		

IPC

	Baseline	D25	D50	D75	D100	I25	I50	I75	I100	P25	P50	P75	P100	P150	MOD
T1	48	52	59	71 .001	73	55	65 .014	59	75	62 .044	59	65	71	82	68 .003
T2	34	46	58 .001	73	78	36	46	60 <.001	65	35	56 .001	59	71	81	69 <.001
T3	56	74 .007	78	86	91	59	80 <.001	82	86	60	78 .001	83	84	91	84 <.001

Experiment 2

AMP

	Baseline	D25& I25	D25& I50	D50& I25	P75& I25	P75& I50	P100& I25	D25& P75	D25&P100	D50&P75	D25, I25& P75
T1	77	80	82	86	81	81	86	80	83	87	87
T2	74	83	86 .032	87	96	95	93 <.001	93	97	91 .001	92 <.001
T3	78	89 .034	92	95	95 <.001	95	96	95 .001	95	98	95 .001

IPC

	Baseline	D25& I25	D25& I50	D50& I25	P75& I25	P75& I50	P100& I25	D25& P75	D25&P100	D50&P75	D25, I25& P75
T1	48	51	53	61	69 .002	71	71	65 .014	72	72	69 .002
T2	34	58 .001	63 <.001	71	78 <.001	80	86	80 <.001	88	88	79 <.001
T3	56	69	85 <.001	90	90 <.001	92	89	95 <.001	95	97	93 <.001

	Baseline	MOD & D25	MOD & D50	MOD & D75	MOD & D100	MOD & I25	MOD & I50	MOD & I75	MOD & I100
T1	48	61	75 <.001	90	90	72 .001	78	78	86
T2	34	84 <.001	92	93	96	76 <.001	80	91	85
T3	56	92 <.001	95	97	98	88 <.001	95	98	99

Table 3: Listener Results for Baseline Condition and All Manipulations Showing Percentage of Utterances Perceived with the Correct Stress Pattern: T1-T3 represent initial,

medial and final target positions, AMP = amplification, IPC = incorrect pitch contour, MOD = modification. Amplification of parameters is indicated as D (duration), I (intensity)

and P (F0) and the percentage of increase. P-values are supplied only for the smallest percentage that reached significance at 5%. All values that are significantly higher than

baseline have been marked in bold.