IS LANGUAGE A FACTOR IN THE PERCEPTION OF FOREIGN ACCENT SYNDROME?

Linda Jose, University of Newcastle, GB

Jennifer Read, Manchester Metropolitan University, GB

Nick Miller, University of Newcastle, GB

Acknowledgements

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. No writing assistance was employed.

Address for correspondence:

Nick Miller

Institute of Health and Society, Speech and Language Sciences,

George VI Building

Newcastle University

Newcastle-Tyne NE1 7RU

Great Britain

Telephone: +44 191 208 56 03

Email: Nicholas.Miller@ncl.ac.uk>

IS LANGUAGE A FACTOR IN THE PERCEPTION OF FOREIGN ACCENT SYNDROME?

ABSTRACT

Neurogenic foreign accent syndrome (FAS) is diagnosed when listeners perceive speech associated with motor speech impairments as foreign rather than disordered. Speakers with FAS typically have aphasia. It remains unclear how far language changes might contribute to perception of FAS independent of accent. Judges with and without training in language analysis rated orthographic transcriptions of speech from people with FAS, speech-language disorder and no FAS, foreign accent without neurological impairment and healthy controls on scales of foreignness, normalness, disorderedness and foreignness.

Control speakers were judged as significantly more normal, less disordered and less foreign than other groups. FAS speakers' transcriptions consistently profiled most closely to those of foreign speakers and significantly different to speakers with speech-language disorder. On normalness and foreignness ratings there were no significant differences between foreign and FAS speakers. For disorderedness, FAS participants fell midway between foreign speakers and those with speech-language impairment only. Slower rate, more hesitations, pauses within and between utterances influenced judgments, delineating control scripts from others. Word level syntactic and morphological deviations and reduced syntactic and semantic repertoire linked

strongly with foreignness perceptions. Greater disordered ratings related to word fragments, poorly intelligible grammatical structures and inappropriate word selection. Language changes influence foreignness perception. Clinical and theoretical issues are addressed.

KEYWORDS: Foreign accent syndrome; motor speech disorders; aphasia; accent perception

INTRODUCTION

Foreign accent syndrome (FAS) represents an acquired speech disorder where a native speaker is perceived as 'foreign', or to speak their native language with a different regional accent, even though they have hitherto spoken with the local accent, do not speak the 'foreign' language of the accent they are perceived to have, have never lived in that country (or other region of their own country) and have no personal associations with that accent (Ryalls & Miller 2014). Broadly the aetiology of FAS can be divided into neurogenic, where the underlying impairment relates to changes brought about by a neurological condition (e.g. stroke, head injury) versus psychogenic, where the origins are associated with behavioural rather than physical or structural changes (Reeves, Burke, & Parker, 2007; Ryalls & Miller 2014). In practice there can be considerable overlap of factors. This article concerns only neurogenic FAS.

In terms of underlying speech pathology FAS is typically linked to an acquired motor speech disorder - apraxia of speech, dysarthria or dysprosody (Katz, Garst, & Levitt, 2008; Kuschmann, Lowit, Miller, & Mennen, 2012; Marien & Verhoeven, 2007; Miller & Lowit, 2014; Miller, Lowit, & O'Sullivan, 2006). These in turn are associated with lesions in left (parietal, frontal, insular) and occasionally right cortex or subcortical networks. The motor speech disorder can result in impairment of

suprasegmental features such as word and sentence stress, tone and rhythm, all of which can influence grammatical and affective output. Segmental production can be altered, producing changes to vowel space, elision of unstressed syllables or consonants from clusters and difficulties pronouncing individual sounds or transitions between (certain) sounds.

If FAS is a manifestation of common underlying motor speech impairments, the question arises of why everyone who acquires these disorders does not sound foreign. One explanation for this (though concord is not unanimous) (Blumstein & Kurowski, 2006; Di Dio, Schulz, & Gurd, 2006) points to a key component of FAS stemming not just from the changes in articulation by the speaker, but crucially, how these changes are perceived in the ear of the listener (Miller et al., 2006; Ryalls & Miller 2014; Van Borsel, Janssens, & Santens, 2005). Where the constellation of articulatory changes most prominent to a listener are linked to natural languages and are evocative of a particular accent with which they are familiar, then a foreign or different regional accent will be perceived. If distortions that are associated with the listener's experience of disordered speech predominate, then a disorder will be heard. A corollary of this means that people perceived to have FAS will commonly have only mild articulation changes, given that more severe and non-natural distortions of segmental and suprasegmental features push the listener's perception into one of

disorder rather than geographical difference. This view is bolstered by the fact that rather than a definite national or regional accent, FAS speakers present with a 'generic foreign accent'. Different listeners hearing the same speaker may perceive different accents; one may sense that the accent alters from day to day (Verhoeven, De Pauw, Pettinato, et al. 2013). Further, certain speech disorders such as stuttering or parkinsonian speech are never reported as sounding foreign. It seems even mild forms of these impairments with their dystonic blocks and prolongations or atonal, asthenic voice never sound natural to listeners.

The descriptive and explanatory literature to date on FAS has shown an almost exclusive focus on articulation. Yet, generally in cases of FAS the presence of (mild) aphasia can also be detected. Furthermore, even when people with FAS perform within normal limits on aphasia tests, they recount problems formulating sentences and difficulty finding words. Characteristically such problems are associated with prolonged pauses, sentence fragments, circumlocution and sometimes semantic paraphasic productions. An aphasic component may lead to alterations in morphological accuracy around e.g. number or tense marking. The accompanying motor speech disorder may also lead to apparent morphological slips due to elision of consonants in clusters or difficulties with specific sounds or consonant transitions. Thus, for instance, for articulatory rather than morphosyntactic reasons, 'he helped

me' may be heard/spoken as 'he help me'; 'the dogs are running' as 'the dog are run/runny'; 'they run fast' as 'they runned fast'. More subtly, an underlying aphasic impairment may be manifest in overall reduced mean length of utterance and restrictions in syntactic complexity and variety.

Reports already hint at the possibility that such features contribute to the perception of foreignness in FAS. Miller et al., (2006) found that both an Italian speaker of English and a speaker with neurogenic FAS produced a number of similar syntactic errors, mostly omissions of grammatical morphemes. Edwards, Patel, & Pople, (2005) reviewed 35 cases of FAS in the literature and found that a third presented with clear agrammatism. Van Borsel et al., (2005) reported prolonged pauses in a FAS speaker's speech and grammatical errors (deletion of articles, improper word order, plural marking slips) and the use of short sentences as factors that judges reported as influencing their perception that the FAS subject was not speaking in their mother tongue. In studies of populations without acquired speech-language disorders it has long been recognised that morphosyntactic or lexical errors can influence accentedness judgments of spontaneous speech (Flege, Munro, & MacKay, 1995).

This study pursues the issue of whether, and if so how far, output changes independent of articulation/ perceived accent might constitute an element in the perception of foreignness in people with neurogenic FAS. This is of interest from the

point of view of the interaction of language and speech factors in the perception of foreignness, from both sociolinguistic and psycholinguistic perspectives – e.g. do language factors play a part; if so, what aspects of language structure and to what degree. From an applied perspective addressing this issue may assist in the diagnosis of what makes foreign accent syndrome foreign, why patients are perceived as foreign, in speakers with FAS; it would highlight language (as opposed to speech) factors that may be manipulated in intervention to help return a speaker to their premorbid perceived native output. Specifically we aimed to examine how native English listeners judged possible differences between written transcriptions from speaker groups when acoustic information was not available. We compared speakers with neurogenic FAS, speakers with speech-language disorder not judged as foreign sounding (SLD), foreign accented speakers (FOR) and native control speakers (CON). We asked:

a) Do raters detect differences between foreign and native speech, 'normal' and disordered speech without the acoustic information?
b) Where are speakers with FAS judged to lie in relation to healthy speakers with a 'true' foreign accent and speakers with motor speech disorders who are not judged as having FAS?
c) If language changes are associated with stronger foreignness and/or

disorderedness judgements, what linguistic features characterise these contrasts?

The intention was to gather broad perceptual impressions from general listeners/ readers, rather than two or three highly specialised linguists. Providing sufficiently informative linguistic analysis, or at least judgements involving metalinguistic awareness, may, however, depend on some degree of formal language analysis training, or at the least general judgements of those with more developed metalinguistic awareness may differ from those with less awareness. Accordingly, we incorporated whether judgements differed between naïve (no knowledge of linguistics, speech language pathology or languages in general) and experienced (judges who do have this knowledge) raters.

Based on auditory perceptual ratings of judges comparing these groups we hypothesized that control speakers would produce output judged as different to all other groups; people with neurogenic FAS would cluster most closely to participants with a true foreign accent; people with aphasia would be rated as different to the speakers with true foreign accent and FAS. We predicted naïve raters would identify fewer elements of output in their evaluation of scripts and show less agreement about which factors might distinguish groups.

METHODOLOGY

Participants

There were two groups of participants: speakers from whom speech samples were gathered and subsequently transcribed and raters who judged these transcripts. All participants were recruited following University Ethics Committee approved procedures. Participation and all activities were conducted following informed voluntary consent, with the right to withdraw without reason and without repercussions at any point in the research.

Speakers

Nineteen female speakers provided spoken output samples that formed the basis of the transcription analyses. Five presented with a neurogenic foreign accent (FAS), diagnosed by an experienced speech and language therapist. They were monolingual native speakers of English with no history of speech-language disorder prior to onset of the altered accent. In common with most people with FAS listener perceived accent was generic and fluctuating (e.g. East European, Mediterranean), but for the present speakers general opinion favoured particular accents (Table 1). The bases of the altered accent included prosodic, rhythmic/syllable structure alterations and segmental distortions. They were selected as being current speech-language clinic patients who

had an altered accent. Presence or not of language disturbance was not a selection criterion.

Table 1 about here

Participants in other groups were chosen to match in gender, general age and educational background. Five participants presented with an acquired speech-language disorder (SLD). All had post-CVA aphasia, classified as mild Broca's type dysfluent aphasia according to the Boston diagnostic classification (Goodglass, Kaplan, Baressi, 2000). Two of this group had an accompanying mild apraxia of speech. Selection for this group included ability to sustain a 30 second monologue, use more than just single word utterances, and that output should not be dominated by apraxic blocks, trial and error run-ups and unintelligible words.

Four participants had true foreign accents (FOR) – French, German, Chinese, Italian. They were adult immigrants to Britain who had learned English to a fluent, proficient standard (retired hospital nurse; retired teacher; secretary; lecturer) but spoke with an agreed perceived foreign accent. They had no history of neurological or speech-language disorders. A control group (CON) consisted of five native English speakers, with no history of neurological illness, speech-language disorder, or foreign language

experience.

Median age for all speakers was 64 years (IQR 60-70; full range 37-81). The people with FAS had median age 62 years, IQR 42-67 (full range 37-70); the SLD group median 67 (IQR 64-71; range 62-72); the FOR group median age 66 (IQR 48-78, full range 45-81); and control speakers median 61 (IQR 55-75, full range 50-81).

Data acquisition

Recordings were made onto digital audio tape using a TASCAM DA-30 recorder with the microphone placed at approximately 45 cm in front of the speaker. Participants described how to make a cup of tea, what they enjoyed as pastimes or recent events. Recordings were transferred to a computer. Two excerpts spliced from each speaker's recording formed the basis for later analyses. These were the first two stretches that represented continuous talk for approximately 30 seconds by the speaker uninterrupted by the interlocutor, regardless of the number of utterances or grammatical status of the content. All samples commenced with the start of an utterance and ended with the termination of an utterance. Splices were matched for total duration and not number of words or syllables. Samples were orthographically transcribed verbatim, including word and sentence fragments, false starts, repetitions, pauses, pause time, fillers (e.g. um, er), as in the following examples:

OK, em :02 If I was making a cup of tea I'd eh :03 put the kettle fi* fill it with water water em and then put the kettle on. And em then I need to get a> depending on making tea tea :05 pot or the mug. 0:31 secs

I had a gentleman who I did the housework for and I had a blind lady who was wonderful. And she eh she didn't get out too much but> She was in apartment and she> we get along very good. We cooked cooked together. And eh lot of lot of nice ladies who who just needed help. 0:32 secs

Data processing

The authors employed the Systematic Analysis of Language Transcripts (SALT; http://www.saltsoftware.com/ accessed 22 December 2014) to examine the characteristics of the passages. The same passages were judged by raters in terms of perceived normalness/ disorderedness/ foreignness. Correlational and regression analyses were conducted to examine the strength of SALT outcomes to predict rating scale evaluations.

For the SALT analysis standard transcription conventions were used to mark unintelligible and abandoned words and utterances. Pauses, omitted words and inflectional morphemes, syntactic and grammatical errors were coded. Utterance level error codes were assigned to clause level errors (e.g. omission or order of clause elements). Word error codes were assigned to phrase level (e.g. omission of articles)

and word level (e.g. the wrong form of an irregular noun) errors. Syntax and morphology were summarised using mean length of utterance (MLU), the number of omissions and syntactic error analysis. The number of different word roots used by each speaker served as a surrogate measure of semantics. Fluency and rate were captured by examining the number and types of mazes (repetitions, revisions, and filled pauses), speaking rate (words per minute) and the number and length of pauses between and within utterances. Measures of the total number of utterances, length of sample and total number of complete words in each speaker's samples were taken. Transcript length was also considered as an influencing factor.

Raters

Sixty-four monolingual, native speakers of English, aged 18-22 years, volunteered as judges (participants were entered into a raffle to win a £10 Sterling, approx. US$15.70 department-store voucher). They were divided into groups designated as experienced-inexperienced. The former (labelled ER) were 38 speech and language therapy students who were, in accordance with previous studies (Southwood & Flege, 1999; Thomson, 1991), fluent in at least one foreign language, had taken speech and language analysis courses and had contact with non-native speakers. The latter (labelled IR) were 26 university students with no experience of disordered or foreign speech, nor of linguistic analysis or rating speech and language performance.

Raters were provided with printed booklets with the 38 extracts (two per speaker) presented in random order. One extract was repeated with a separate code number to test for intra-rater reliability. The booklets contained details of annotations employed and standardised instructions for rating across the three dimensions (below). The exact duration of each passage was provided for raters alongside the transcription in order to give an impression of the overall utterance rate. Raters were unaware of the purpose of the study or the identity of the speakers, or that there were subgroups. They completed the booklets in their own time without conferring, and as such were able to (re)read samples as often as they required to reach a judgement. Approximate time to complete was 40 minutes.

Raters were randomly assigned to one of three rating groups. Group 1 rated how 'normal' they judged the output to be ('normalness'). Group 2 rated the samples in terms of likelihood that the speaker was foreign ('foreignness'). Group 3 evaluated samples in relation to likelihood that the speaker had some kind of speech-language disorder ('disorderedness'). Normalness and disorderedness were each rated by two groups of 13 experienced and 9 inexperienced judges (44 total); foreignness by 12 experienced and 8 inexperienced (20 judges).

Ratings entailed a 7 point equal appearing interval scale ranging from 1 ‘definitely normal/local’ to 7 definitely not normal/definitely foreign/definitely disordered. Our main focus was the foreignness scale. The other scales were included to control for the possibility that raters might judge the FAS and FOR groups as not normal simply on the basis of being different. Hence the normalness scale. We also wished to know whether there might be features that were regarded as different but disordered rather than foreign and whether this is what listeners/readers perceive when they judge as foreign vs disordered. Hence the disorderedness scale.

Additionally, raters were asked to underline in the samples what had influenced their judgment and provide a free text explanation. These were utilised to aid interpretation of what elements of production raters might associate with perceptions of foreignness and disorderedness. These free text comments and underlinings were used to by the authors to arrive at a list of all possible factors perceived by raters. The ten most common features mentioned across all raters and samples were extracted for further analysis (see results). A score of one was attributed to the relevant category if a rater indicated that this influenced their rating of a particular speech sample, irrespective of how many instances of that variable appeared in the sample. These were added together, giving a total score for each speaker in each of the ten categories.

Data processing

The mean score for every sample was calculated for both experienced and inexperienced raters on each of the three dimensions. A 2x4 independent measures ANOVA with Bonferroni post hoc analyses was performed for each of the three rating dimensions to determine whether there were any statistically significant differences in ratings between the separate speaker groups (FAS, SLD, FOR, CON) and whether there were any significant differences between the experienced and inexperienced raters.

Spearman's correlations were performed to investigate which SALT linguistic and perceived output features were most strongly associated with perceptions of foreignness, disorderedness and normalness. Backwards linear regression was conducted on factors that showed moderate to high correlations ($r > 0.4$) to establish which factors best accounted for the variance in the ratings.

Reliability

Inter-rater reliability of the orthographic SALT transcription was measured by a judge who was blind to the nature of the study, coding four randomly selected transcripts, one from each of the speaker groups and comparing percentage agreement with the first coder for each of the linguistic analyses. Inter-rater reliability for categorising and coding the qualitative data was carried out based on

two booklets from each of the three rater groups. One sample was repeated with a different code in the booklet to examine intra-rater reliability.

From 19 categories of linguistic measures computed in SALT there was 100% agreement in all categories for one speaker. For the other three there was 100% agreement on measures apart from number of complete words, where agreement was 97.5%. There was a difference of five words across all the transcripts. This in turn slightly altered measures of MLU and words per minute but not statistically significantly.

Taking the scores from all raters (n 51) who had not indicated that the repeated sample inserted for intra-rater agreement purposes was one they had already evaluated, there was a significant correlation between the first and second rating scale scores ($r_S = 0.775$, $p<.001$) and no significant difference between the scores awarded first or second time. Looking at experienced and inexperienced raters separately correlations and differences were: experienced raters $r_S = 0.821$, $p <.001$, $z = 0.000$, $p = 1.000$; inexperienced raters $r_S = 0.667$, $p <.001$, $z = 0.004$, $p = 1.000$. Examination of the number and type of reasons given for their judgements across the repeated samples revealed significant correlations and no significant differences for the whole rater group. The same applied to the experienced group alone. However, for the inexperienced raters the correlation between factors was not significant (r .149, p

0.53), despite there being no significant differences in total items.

RESULTS

Transcript ratings

The first question concerned whether raters perceived any differences between speaker groups based on the different rating scales. Table 2 summarises the descriptive statistics for rating scale judgements by speaker and rater group for each of the three dimensions (the varying n relates to the differing number of raters per group and fewer FOR speakers).

Table 2 about here

A 2 (rater groups) x 4 (speaker groups) independent subjects ANOVA with Bonferroni post hoc tests was conducted for each of the three rating dimensions to examine for possible differences between rater and speaker groups and any possible interactions.

'Foreignness' ratings

There was a highly significant main effect of speaker group ($F(3, 372) = 55.30$, $p < .001$) and an interaction between rater group and speaker group, $F(3, 372) = 3.760$, $p = 0.011$. Bonferroni post hoc tests revealed that foreignness ratings attributed to the control group were significantly lower (more 'native') compared to each of the other three speaker groups ($p<.001$). Both the experienced and inexperienced listeners rated

the control speakers as the least foreign and the FAS and foreign speakers very similarly and higher on degree of foreignness. There was a just significant main effect of rater group on the foreignness rating of speakers with SLD ($F(1, 372) = 4.170$, $p = 0.042$). The SLD group were judged as more foreign by the inexperienced raters (mean rating 4.69, SD 1.35) than the experienced raters (mean rating 3.57, SD 1.59). The inexperienced raters scored SLD speakers highest for foreignness, whilst experienced raters attributed higher foreignness ratings to those with foreign accent syndrome, although not statistically higher than the foreign and SLD speakers. There were no significant differences between the ratings given to the other speaker groups.

'Normalness' ratings

There was a highly significant main effect of speaker group ($F(3,405) = 115.477$, $p < .001$) but no interaction between rater group (experienced vs inexperienced) and speaker group. Bonferroni post hoc tests revealed no significant difference between ratings of FOR and FAS speakers ($p = 0.528$). Contrasts between all other speaker groups were individually highly significant ($p<.001$). The control group were judged to be significantly more 'normal' than both the FOR and FAS groups, and the SLD group to be the least normal. There was no significant main effect of rater group on the normalness dimension ($F(1, 405) = 0.345$, $p=0.557$).

'Disorderedness' ratings

There was a significant main effect of speaker group ($F (3, 410) = 97.318$, $p < .001$) but no significant interaction between rater and speaker groups ($F (3,410) = 1.559$, $p = 0.199$). There was no significant main effect of rater experience on judgments of disorderedness ($F (1, 410) = .145$, $p = 0.703$). Bonferroni post hoc tests revealed that differences in ratings between each of the speaker groups were individually highly significant. The most significant contrast was between the control group and all the other speaker groups ($p < .001$), followed by the contrast between the SLD and FAS group ($p=0.009$) and then the FAS and FOR group ($p=0.018$). The SLD group was rated as most disordered and the control group as least disordered. FAS speakers were rated most similarly to foreign speakers for degree of disorderedness.

Perceived influential factors

Each rater indicated through underlining and free text which aspects of the passages influenced their judgements. Responses were categorised into the ten most common features identified across all ratings. This yielded: speech rate, abandoned words and utterances, between-utterance pauses, within-utterance pauses, filled pauses, repetitions, omitted words, omitted inflectional morphemes, phrase and word level syntactic errors, and clause and utterance level syntactic errors. Correlation and regression analyses were performed to explore which factors were most strongly associated with ratings of normalness, foreignness and disorderedness. Except where

indicated, tests were conducted using a two-tailed alpha level of .05 and only significant outcomes are discussed.

'Foreignness'

A weak significant correlation existed between the total feature categories and foreignness ratings (Spearman's r = 0.372, p = 0.021). At an individual category level, word level syntactic errors had the strongest relationship with foreignness rating (r = 0.519, p = 0.001), followed by omission of morphemes (r = 0.487, p = 0.002). Significant but weak relationships were present for within utterance pauses (r = 0.366, p = 0.024) and between utterance pauses (r = 0.345, p = 0.034). Using a regression model to explore further, the number of omitted morphemes and within utterance pauses were the best predictors of foreignness (R = .487), accounting for around 24% of the variance. The perceived number of omitted morphemes was the only variable with a significant independent effect, indicating this as having the greatest influence on foreignness ratings.

'Normalness'

The sum of the ten factors correlated with the perception of normalness for each speaker (r = 0.671, p <.001). Three factors had an independent, highly significant relationship with the rating score: speech rate (r = 0.557, p <.001), between utterance pauses (r = 0.641, p <.001) and within utterance pauses (r = 0.643, p <.001).

Backwards multiple linear regression showed that speech rate and within utterance pauses accounted for approximately 39% of the variance ($R = 0.621$, $p<.001$) and combining these variables with between utterance pauses did not account for any more of the variance. No factors within the model had a significant independent effect when the other variables were held constant.

'*Disorderedness'*

The sum of all reported influential factors had a strong and highly significant relationship with the disorderedness rating ($r = 0.795$, $p <.001$). Between utterance and within utterance pauses had positive, highly significant relationships ($r = 0.669$, $p <.001$; $r = 0.619$, $p<.001$). Speech rate ($r = 0.448$, $p = 0.005$), number of abandoned words and utterances ($r = 0.376$, $p = 0.020$) and syntactic word level errors ($r = 0.362$, $p = 0.026$) had weak significant correlations. Filled pauses had a weak negative but significant relationship ($r = -0.366$, $p = 0.026$). Regression modelling using these significant factors showed that abandoned words and utterances, between utterance pauses, within utterance pauses and word level syntactic errors had the greatest influence on raters' perception of disorderedness and together accounted for 68% of the variance ($R = 0.826$, $p<.001$). Word level syntactic errors was the only variable with a significant independent effect, suggesting this had the greatest influence on raters perceptions of disorderedness.

Linguistic factors

Based on the SALT coding of samples a further set of analyses examined the strength of association of these variables with perceived normalness/ disorderedness and foreignness ratings. There was a strong positive correlation between the total number of linguistic 'errors' and normalness ratings ($r = 0.562$, $p < .001$); foreignness ratings ($r = 0.567$, $p < .001$); and perceived disorderedness ($r = 0.526$, $p = 0.001$). Results for the individual features that proved significant in association with the different rating scales appear in tables 3-54 (see Appendices). All non-significant relationships are omitted.

Tables 3-5

Rating scales were not divided by unique sets of SALT categories showing a significant correlation with a particular dimension. Normalness, disorderedness and foreignness all showed instances of changes to all the SALT categories, aside from omitted words associating significantly with judgements of foreignness but not normalness/ disorderedness, and sample length and number of utterances relating significantly to normalness and disorderedness but not foreignness. However, the relative prominence of the different SALT categories differed between rating scales. As predicted, given that they are different sides of the same coin, rank order of strength of associations across

the categories correlated highly significantly for normalness and disorderedness ratings (Spearman's r .825, p = <0.001). However, rank order of strength of association of categories associated with foreignness did not correlate significantly with either normalness (r -.092, NS) or disorderedness (r .123, NS).

The likelihood existed that several SALT categories are closely related and do not exert an independent influence on judgements. To establish which factors might have an independent effect when interrelationships were controlled for, we carried out backward regression analyses entering the features that attained significant correlations with rating scale outcomes (tables 6-7).

With respect to perceived foreignness table 6 indicates the most favourable combination of factors from table 3 for prediction of foreignness rating.

Table 6 about here

These variables account for approximately 74% of the variance (R=0.862). Entering all other factors in table 4 accounted for only another maximum 5% variance. All have independent effects with foreignness rating apart from repetition, with word level syntactic errors being the strongest independent predictor.

As regards normalness, regression modelling selected the factors in Table 7, which together accounted for approximately 97% of the variance (R = .982).

Table 7 about here

Between utterance pause duration was excluded as having no effect on variance, whilst all factors except repetitions had independent effects. Thus, changes in occurrence of any one of these linguistic variables when other variables are held constant would increase the likelihood a sample would be considered more/less normal; an increase in repetitions without a simultaneous increase in the other variables would not.

Table 8 gives the factors from table 4 with the strongest prediction of disorderedness rating. Together they account for approximately 96% of the variance (R = 0.977).

Table 8 about here

The factors have independent effects (except the number of words per minute), meaning an increase in any of these factors when others are held constant increases the likelihood of the speech being rated as more/less disordered.

DISCUSSION

Results show that both naïve and experienced listeners judge transcriptions from the non-control speakers as significantly different from the control group, whether judgements relate to normalness, disorderedness or foreignness. The scripts from speakers with FAS were consistently rated most closely to those of foreign speakers and significantly different to speakers with a SLD, who were generally judged to be furthest from normal. On ratings for normalness and foreignness there was no significant difference between FOR and FAS speaker scripts. For disorderedness, although in rank order people with FAS were still judged as closest to FOR, they actually fell midway between FOR speakers and those with a SLD only, with statistically significant differences in either direction. This picture confirms hypotheses.

Contrary to hypotheses, apart from a borderline, and possibly chance (given the number of analyses conducted), difference between naïve and experienced raters on foreignness rating, (naïve raters scored SLD scripts as sounding more foreign than the experienced raters; experienced raters had the FOR group as most foreign, but not significantly), familiarity with grammatical analysis and speech evaluation had no significant bearing on outcomes. This suggests that on this task heightening of metalinguistic awareness through training in language analysis and speech pathology did not confer any advantage in terms of sensitivity to departures from expected

language structure. The changes that mark speakers as foreign or disordered were apparent to 'untrained' individuals, and these raters did not arrive at significantly different sets of linguistic features attracting their attention compared to the experienced raters. Whilst this runs contrary to some previous research that suggests trained raters may behave differently in rating such variables as voice quality or nasality, on the other hand it is not an unexpected finding. One does not need to be a trained phonetician to detect and describe a different accent. Specialist training may lend an advantage when it comes to offering explanations for changes, but this was not the task here.

These findings confirm that even without auditory information, the spoken output of people with FAS is characterised by features that mark them out as neither native nor disordered and more likely as foreign. The speakers with SLD display characteristics that separate them out as different to all others. Notably, when raters are requested to base judgements on normalcy and foreignness speakers with FAS clearly cluster alongside the FOR group – there are aspects of their output judged as different to normal, but not so strongly that it evokes perceptions of disorder. However, when raters are asked to focus on disorderedness there are clearly some aspects of FAS output that do suggest to raters the individuals are not straightforwardly just foreign. These results echo findings from auditory perceptual examinations, that conclude

people with FAS are more likely to be perceived as foreign rather than disordered, even when their speech does contain clear elements of disordered pronunciation (Dankovičová & Hunt, 2011; Miller et al., 2006).

Rater judgements on what they felt influenced their decisions and the SALT linguistic analyses point to some key commonalities concerning which variables play a part. Rate of output and the hesitancies and pauses within and between utterances clearly played a role in delineating control scripts from others. A range of linguistic features also figure prominently in swaying judgements - most conspicuously, word level syntactic and morphological deviations and reduced syntactic (e.g. mean length of utterance) and semantic (e.g. limited word root variation) repertoire.

A key question entails whether there are particular deviations that mark people as foreign versus disordered and does this possible division account for the position of the FAS group in relation to the FOR and SLD groups. On the one hand the multiplicity of factors common across all judgement tasks and non-control groups militates against such a view. There is a cumulative rather than feature-specific effect; the more deviant linguistic features, the more a speaker is perceived to be 'non-normal'. Indeed, for non- normal and disordered ratings, the sum of factors account for approaching 100% of the variance, though for foreignness perception it

was only around 74%, indicating that there are other and possibly more subtle features that readers are picking up and that foreignness factors are possibly judged as closer to 'normal'.

Nevertheless, there did exist some variation in the strength of individual features in marking out one dimension or another. Transcriptions judged to be more foreign were characterised by more prominent word level syntactic 'errors' and little variety in word roots in the presence of frequent pauses. Words per minute and presence of incomplete words and particularly long pauses, rather than just total pauses, were more strongly associated with disorderedness impressions. Thus, people with true foreign accents were 'heard' as slower, more hesitant speakers, with difficulty producing word agreement or correct morphological marking and employing predominantly simple grammatical structures. Judgements crossed over to disorderedness when word fragments, poorly intelligible grammatical structures and inappropriate word selection were to the fore. Arguably, when readers were primed to rate according to their views of foreignness, their attention was drawn to the features which emphasized the commonalities with true foreign speakers. However, when the focus was on features associated with disorderedness the paucity of these in the FOR groups scripts caused the ratings of the FAS speakers to shift from FOR towards SLD. Several important conclusions can be drawn from these observations.

Even when people cannot hear people with FAS, there is a quality to their output that evokes in raters the impression of foreignness but not disorderedness. Within the conceptualisation of FAS as in the ear-of-the-listener as much as the mouth-of-the-speaker, the perception of a different accent versus disorder is probabilistic in nature. Speakers with FAS, if one analyses their speech acoustically or with narrow phonetic transcription, demonstrate changes in their speech that are found both in other natural languages and derailments that are found only in disordered populations (Dankovičová & Hunt, 2011; Miller et al., 2006; Roy, Macoir, Martel-Sauvageau, & Boudreault, 2012). FAS is assigned by listeners when speech production reminiscent of a given accent is to the fore. When speech derailments associated only with disordered speech are uppermost in listeners' perceptions, then a speech-language disorder is labelled.

Reactions, though, are likely not based on an entirely accumulative basis. Isolated or infrequent cues that impinge listeners'/readers' consciousness that are not reconcilable with their first hypotheses regarding the speaker's speech may bring about a categorical shift. This phenomenon is found not just in diagnosis of FAS versus motor speech disorders, but in numerous other sociolinguistic judgments such as social class, gender or age (Llamas & Watt, 2010).

However, although accent may index social group, age and so forth, interlocutors consciously or subconsciously detect other language and non-language cues to arrive at their categorisation, such as word choice, variety of syntactic structures, physical appearance. The current findings suggest that language output is one of these additional features that can exercise an influence on perceptions of foreignness versus disorderedness in people with FAS. While output is in keeping with the profile of production of foreign speakers, people with FAS are viewed as foreign. Even when they cannot be heard, the presence of language features associated with disordered rather than foreign speech can bias reactions away from foreignness towards disorderedness.

The transcriptions read by the different rater groups were the same scripts. If merely the presence of features of disordered output determined evaluations one would have expected the FAS group to have associated most closely with the SLD group on all occasions. However, it was only when readers were primed by the request to monitor for disorderedness that the presence of features of language impairment crossed a threshold to be taken into consideration.

The present work constituted an exploratory study to establish whether there is evidence that language factors may serve a role in foreignness judgements (of people with FAS) in addition to speech signal perception, and to gain preliminary insights into what these

factors might be. The findings suggest that indeed language factors do play a part. Nevertheless, there exist some provisos against which outcomes here have to be weighed.

Firstly, different listener groups rated the separate dimensions, based on the supposition that judgements would be contaminated across tasks if raters saw they were evaluating the same scripts. There is therefore a risk that differences relate to rater and not rating scale variation. Ideally a follow-up study might employ an alternative design that, for instance, randomly assigns scripts, and on what dimension they are to be rated, across one single rater group. However, there are indications from the data that this was not a major confound here.

The two rating scales one would expect to be closely related to each other as different sides of the same coin, i.e. normalness vs disorderedness, correlated highly significantly with each other in terms of rank order of prominence of features associated with the dimensions identified by different rater groups. The foreignness features identified and their rank orders differed significantly from normalness and disorderedness ratings. This pattern of rank order similarities but rating scale differences suggests the results do not arise from rater contrasts as opposed to speaker contrasts.

Further, we anticipated naïve vs experienced judges would identify different variables

and rate dimensions differently and so analysed their responses separately. This difference did not transpire. To a degree one might therefore take this as a partial replication of ratings with a different group of judges, though it does not preclude a separate study to definitively confirm the supposition.

Secondly, whilst transcriptions removed all clues to accent, they did not remove clues to other indicators of output such as pause length and duration of the sample. The visual appearance of transcripts may have swayed interpretations. Based on this one might expect judgements to be dominated by these clues. Pause behaviours did feature as influencing variables. On the one hand this is expected, given that pauses are known to represent a powerful indicator of speaker status. The fact that raters identified word, morphological and sentence level factors as significant suggests, though, that their evaluations were not unduly biased by visual appearance nor based solely on pause behaviours. However, again, to settle the issue a future study might employ transcriptions with purely word and sentence level detail included to examine whether similar results obtain.

A further possible issue entails whether a perceptual examination of coarse distinctions as applied here is capable of detecting potential divergences between speaker groups and whether the question posed in the study should rather be

addressed exclusively through fine-grained linguistic analysis? The two perspectives, however, answer different questions and are not mutually exclusive. The focus here was on whether members of the general public do perceive any differences between speaker groups, and if so, give an indication of what factors they may be using in their judgements. Speakers may differ from 'normal' on a whole range of features that could emerge from a fine-grained analysis, but this would not deliver an answer to which ones are prominent markers in everyday social encounters. A similar contrast pertains in phonetic comparisons between FAS speakers and other groups. Acoustic, kinematic or articulatory phonetic examinations provide a detailed description of what speakers may be doing, but they do not of themselves say why listeners hear one person as foreign and the other not. Acoustic and articulatory phonetic descriptions of FAS speech find the same changes that are present in the speech of people with motor speech disorders who are not perceived as foreign or who are heard as disordered (Dankovičová & Hunt, 2011; Miller et al., 2006; Roy, Macoir, et al., 2012). The fact that there were next to no significant differences between rater groups and their identified features suggests more vs less expert judges 'hear' the same features when making judgements. The fact that rater derived categories of difference largely reflect SALT analyses seems to indicate that the present arrangement of raters and what they were instructed to do reliably served the purpose of the study. To elucidate the relationship between these two perspectives a

future investigation might apply detailed linguistic analysis to transcripts followed by some form of factor/ discriminant analysis to ascertain whether variables shown to discriminate between groups are also the factors perceived by judges to distinguish them.

A final issue concerns the speaker groups. Although 19 speakers provided data, the groups here were relatively small. There is scope for extension of group sizes, not just to expand numbers, but to capture more the heterogeneity that can exist within groups (e.g. different types and severity of aphasia). The presence of more overtly aphasic speakers and/or less proficient foreign speakers for instance may also aid in highlighting the types of language impairment that are associated exclusively with disorder rather than foreignness. It may also indicate what elements of fluency or morphosyntactic departure would never place a speaker in the foreign or control group. Similarly more extensive data per speaker may enable refinement of precisely what kinds of pauses, morphological or syntactic deviations prompt assignment to which category.

This was an exploratory study to ascertain if judges do perceive groups to differ on language features, if so which. The outcomes suggest that they do and some indications for more prominent features have been delivered. These can be used to guide more in-

depth and systematic follow-up studies to confirm if findings are upheld in other circumstances and to explore with larger speaker groups. Further, the data from patients came from FAS and language impairment associated with neurological aetiologies. Studies have examined in what ways acoustic or speech perceptual analyses might support differential diagnosis between neurological and psychological origins. The present study offers a method for investigating whether listener/reader perceived differences may also be reliable in dividing off such alternative aetiologies.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors

REFERENCES

Blumstein, S. E., & Kurowski, K. (2006). The foreign accent syndrome: A perspective. *Journal of Neurolinguistics, 19*(5), 346-355.

Dankovičová, J., & Hunt, C. (2011). Perception of foreign accent syndrome speech and its relation to segmental characteristics. *Clinical Linguistics & Phonetics, 25*(2), 85-120.

Di Dio, C., Schulz, J., & Gurd, J. (2006). Foreign Accent Syndrome: In the ear of the beholder? *Aphasiology, 20*(9 - 11), 951-962.

Edwards, R. J., Patel, N. K., & Pople, I. K. (2005). Foreign accent following brain injury: Syndrome or epiphenomenon? *European Neurology, 53*(2), 87-91.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America, 97*(5 Pt 1), 3125-3134.

Goodglass, H., Kaplan, E., & Baressi, B. (2000). *The Boston Diagnostic Aphasia Examination*. Philadelphia: Lippincott, Williams and Wilkins.

Katz, W. F., Garst, D. M., & Levitt, J. (2008). The role of prosody in a case of foreign accent syndrome. *Clinical Linguistics & Phonetics, 22*(7), 537 - 566.

Kuschmann, A., Lowit, A., Miller, N., & Mennen, I. (2012). Intonation in neurogenic foreign accent syndrome. *Journal of Communication Disorders, 45*(1), 1-11.

Llamas, C., & Watt, D. (Eds.). (2010). *Language and identities*. Edinburgh: Edinburgh University Press.

Marien, P., & Verhoeven, J. (2007). Cerebellar involvement in motor speech planning: Some further evidence from foreign accent syndrome. *Folia Phoniatrica et Logopaedica, 59*(4), 210-217.

Miller, N., & Lowit, A. (Eds.). (2014). *Motor speech disorders: a cross language perspective*. Clevedon: Multilingual Matters.

Miller, N., Lowit, A., & O'Sullivan, H. (2006). What makes acquired foreign accent syndrome foreign? *Journal of Neurolinguistics, 19*(5), 385-409.

Reeves, R. R., Burke, R. S., & Parker, J. D. (2007). Characteristics of psychotic patients with foreign accent syndrome. *J Neuropsychiatry Clin Neurosci, 19*(1), 70-76.

Roy, J.-P., Macoir, J., Martel-Sauvageau, V., & Boudreault, C.-A. (2012). Two French-speaking cases of foreign accent syndrome: an acoustic–phonetic analysis. *Clinical Linguistics & Phonetics, 26*(11-12), 934-945.

Ryalls, J., & Miller, N. (2014). *Foreign accent syndromes: the stories people have to tell*. New York and Oxford: Guilford Press.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics, 13*(5), 335-349.

Thomson, I. (1991). Foreign accents revisited: the English pronunciation of Russian

immigrants. *Language Learning, 41*(2), 177-204.

Van Borsel, J., Janssens, L., & Santens, P. (2005). Foreign accent syndrome: An organic disorder? *Journal of communication disorders, 38*(6), 421-429.

Verhoeven, J., De Pauw, G., Pettinato, M., Hirson, A., Van Borsel, J., & Mariën, P. (2013). Accent attribution in speakers with Foreign Accent Syndrome. *Journal of Communication Disorders, 46*(2), 156-168.

Speaker	Age	Perceived accent	Medical Diagnosis
FAS 1	70	French	CVA (middle cerebral artery)
FAS 2	64	Italian	Subarachnoid haemorrhage (right frontol medial)
FAS 3	47	German/Polish	Vasculitis
FAS 4	62	French	CVA (middle cerebral artery)
FAS 5	37	Asian	Paediatric autoimmune neuropsychiatric disorder (PANDA) related action dystonia.

Table 1. Participant information (FAS group).

Speaker Group	Rater Group (foreignness)	M	SD	n
FAS	ER	3.98	1.36	60
	IR	4.1	1.61	40
	Total	4.03	1.46	100
FOR	ER	3.78	1.49	48
	IR	3.83	1.58	32
	Total	3.8	1.52	80
SLD	ER	3.57	1.59	60
	IR	4.69	1.35	40
	Total	4.02	1.59	100
CON	ER	1.93	1.06	60
	IR	1.84	1.04	40
	Total	1.90	1.05	100
Speaker Group	Rater Group (disorderedness)	M	SD	n
FAS	ER	4.67	1.46	65
	IR	5.06	1.32	45
	Total	4.83	1.41	110
FOR	ER	4.22	1.61	52
	IR	4.24	1.22	36
	Total	4.23	1.46	88
SLD	ER	5.60	1.38	65
	IR	5.19	1.58	45
	Total	5.43	1.47	110
CON	ER	2.42	1.42	65
	IR	2.21	1.06	45
	Total	2.33	1.29	110
Speaker Group	Rater Group (normalness)	M	SD	n
FAS	ER	4.56	1.68	65
	IR	4.34	1.37	45
	Total	4.47	1.56	110
FOR	ER	4.17	1.36	52
	IR	4.10	1.26	36
	Total	4.14	1.31	88
SLD	ER	5.66	1.38	64
	IR	5.46	1.30	45

	Total	5.58	1.35	109
CON	ER	2.04	1.15	61
	IR	2.22	1.18	45
	Total	2.12	1.16	106

Table 2. Descriptive Statistics for Ratings By Rater and Speaker Groups For Each Rating Dimension (M = Mean; SD Standard Deviation; ER = Experienced Rater Group; IR= Inexperienced).

Linguistic factor	r	p	Relationship
Number of word level syntactic errors	.556	<.001	Strong positive
Number of different word roots	-.523	.001	Strong negative
Between utterance pause duration	.481	.002	Weak positive
Number of repetitions	.477	.002	Weak positive
Number of filled pauses	.471	.003	Weak positive
Number of complete words	-.456	.004	Weak negative
MLU in morphemes	-.452	.004	Weak negative
Words per minute	-.450	.005	Weak negative
MLU in words	-.432	.007	Weak negative
Number of omitted words	.427	.007	Weak positive
Number of between utterance pauses	.377	.020	Weak positive
Within utterance pause time	.374	.021	Weak positive
Number of within utterance pauses	.363	.026	Weak positive

Table 3 Significant Correlations of Variables with Foreignness Ratings and Coefficients of Strongest Linguistic Predictors of Foreignness.

Linguistic factor	r	p	Relationship
Words per minute	-.730	<.001	Strong negative
Number of different word roots	-.719	<.001	Strong negative
Number of complete words	.713	<.001	Strong positive
Number of within utterance pauses	.672	<.001	Strong positive
Between utterance pause duration	.671	<.001	Strong positive
Within utterance pause time	.671	<.001	Strong positive
Total between utterance pauses	.635	<.001	Strong positive
Number of repetitions	.547	<.001	Strong positive
MLU in morphemes	.540	<.001	Strong positive
MLU in words	.502	<.001	Strong positive
Number of utterances	-.487	.002	Weak negative
Sample length (secs)	-.462	.004	Weak negative
Number of filled pauses	-.445	.005	Weak negative
Number of word level syntactic errors	.369	.023	Weak positive

Table 4. Significant Correlations of Variables with Normalness Ratings and Coefficients of Strongest Linguistic Predictors of Normalness.

Linguistic factor	r	p	Relationship
Between utterance pause duration	.667	<.001	Strong positive
Words per minute	-.667	<.001	Strong negative
Number of different word roots	-.631	<.001	Strong negative
Number of complete words	-.600	<.001	Strong negative
Number of within utterance pauses	.583	<.001	Strong positive
Within utterance pause duration	.571	.001	Strong positive
Number of filled pauses	-.524	.001	Strong positive
Number of repetitions	.519	.001	Strong positive
Number of between utterance pauses	.502	.001	Strong positive
Number of word level syntactic errors	-.473	.003	Weak negative
Number of utterances	-.456	.004	Weak negative
MLU in morphemes	-.448	.005	Weak negative
MLU in words	-.407	.011	Weak negative
Total length of sample (secs)	.354	.029	Weak positive

Table 5. Significant correlations of variables with disorderedness ratings and coefficients of strongest linguistic predictors of disorderedness.

Linguistic variable	**β**	**t (df =33)**
Word level syntactic errors	.787	7.666**
No. Complete words	-.372	3.194**
Repetitions	.369	1.871 NS
MLU in morphemes	-.174	2.896**

**Significant at the 0.01 level, NS Not significant

Table 6. Coefficients of linguistic measures calculated in SALT found to be the best predictors of foreignness from backward multiple linear regression.

Linguistic variable	β	t (df = 24)
Word level syntactic errors	1.040	12.840**
Filled pauses	.793	7.943**
Total between utterance pause time	.494	7.209**
Sample length	-.646	-6.339**
Number of complete words	-3.027	-5.877**
Within utterance pauses	1.565	5.624**
Different word roots	2.225	5.486**
MLU in words	-1.149	-3.150**
MLU in morphemes	1.227	2.523*
Total within utterance pause time	-.655	-2.457*
Words/minute	-.279	2.437*
Number of utterances	.895	2.367*
Repetitions	.162	1.983 NS

**Significant at the 0.01 level; * Significant at the 0.05 level; NS Not significant

Table 7. Coefficients of linguistic measures calculated in SALT found to be the best predictors of normalness rating from backward multiple linear regression.

Linguistic variable	β	t (df=25)
Repetitions	.353	6.282**
Sample Length	-.640	-6.070**
Within utterance pauses	1.218	5.899**
No. Complete words	-1.989	-5.534**
No. Different word roots	2.083	4.792**
Between utterance pauses	.338	4.765**
MLU in morphemes	-2.121	-4.709**
MLU in words	1.734	-3.821**
Words/ minute	-.241	-1.923 NS

**significant at the 0.01 level; *significant at the 0.05 level; NS Not significant

Table 8. Coefficients of linguistic measures calculated in SALT found to be the best predictors of disorderedness rating from backward multiple linear regression.