This is a repository copy of *On the potential of phonetic analysis to distinguish between people with epilepsy and non-epileptic seizures*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/148017/

Version: Accepted Version

Article:

Walker, G. orcid.org/0000-0001-5022-4756, Walker, T., Moon, L. et al. (1 more author) (2020) On the potential of phonetic analysis to distinguish between people with epilepsy and non-epileptic seizures. International Journal of Applied Linguistics, 30 (1). pp. 92-109. ISSN 0802-6106

https://doi.org/10.1111/ijal.12268

This is the peer reviewed version of the following article:Walker, G, Walker, T, Moon, L, Reuber, M. On the potential of phonetic analysis to distinguish between people with epilepsy and non-epileptic seizures. Int J Appl Linguist. 2019; 1– 18. , which has been published in final form at https://doi.org/10.1111/ijal.12268. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

This is the author's final version of an article to be published in *International Journal of Applied Linguistics*: https://onlinelibrary.wiley.com/journal/14734192. There may be small differences between this version and the final published version. The final published version should be consulted before quoting or discussing in detail.

On the potential of phonetic analysis to distinguish between people with epilepsy and non-epileptic seizures

Authors: Gareth Walker[a], Traci Walker[b], Lauren Moon[a] and Markus Reuber[c]

[a] School of English, University of Sheffield
[b] Department of Human Communication Sciences, University of Sheffield
[c] Academic Neurology Unit, University of Sheffield

Acknowledgements: This work was supported by the Sheffield Undergradute Research Experience scheme at the University of Sheffield, which allowed the third author to work under the supervision of the first two authors.

Abstract

A body of research has shown that there are linguistic differences in the way people with epilepsy talk about their seizures when compared to those with non-epileptic seizures. We extend this line of research by presenting the results of a phonetic analysis comparing speech samples from people with a confirmed diagnosis of epilepsy (7 patients), to those with a confirmed diagnosis of non-epileptic seizures (8 patients). Variables considered include features of pitch, intensity, duration and pausing in their responses to questions from a neurologist during medical history-taking. We find only limited evidence of differences between the two diagnostic groups (epilepsy vs. non-epileptic seizures). We discuss possible reasons for this lack of evidence.

Keywords: phonetics, epilepsy, non-epileptic seizure

Background

The differentiation between the most common causes of Transient Loss of Consciousness represents a clinical challenge. Tests performed after the event lack sensitivity and specificity (Kotsopoulous et al., 2003). (The sensitivity of a clinical test refers to the test's ability to identify all those with the disease; specificity refers to the test's ability to correctly identify all those without the disease.) The clinician's interpretation of descriptions of the event provided by patients (and witnesses when available) continues to make the most important contribution to the diagnosis (Malmgren, Reuber, & Appleton, 2012). Whereas fainting (syncope) can be distinguished reliably from the other two most important causes (epilepsy and nonepileptic seizures, henceforth EP and NES respectively) on the basis of a modest number of factual enquiries, the distinction of epilepsy from NES is more challenging. In fact, the great majority of patients ultimately diagnosed with NES initially receive an inappropriate diagnosis of (and treatment for) epilepsy, and it typically takes several years before the correct diagnosis of NES is made (Reuber et al. 2002). However, a body of research has shown that there are linguistic differences in the way people with epilepsy talk about their seizures when compared to those with non-epileptic seizures (Plug & Reuber, 2009; Plug, Sharrack, & Reuber, 2009a, 2009b, 2009c). Most notably, Reuber, Monzoni, Sharrack, and Plug (2009) show, in a blinded, multi-rater study, that 85% of ultimate medical diagnoses were correctly predicted using only qualitative linguistic criteria, such as whether a description of subjective seizure symptoms is volunteered (EP) or not provided even when prompted (NES); or whether seizures are repeatedly conceptualised as a fight or struggle (EP) or not conceptualised as such (NES). While the differences described in these studies were identified by careful post-hoc study of transcripts and video-recordings, subsequent work has shown that clinicians are able to alter their history-taking by asking more open questions, and allowing patients more opportunity to describe their problem in their own way, and that these features can make a contribution to the differential diagnosis when doctors simply listen out for them as they talk to patients (Jenkins & Reuber, 2014; Jenkins et al., 2016). The distinction of these two common seizure disorders is of great importance because they have different causes and require different treatments: epilepsy is caused by abnormal electrical discharges in the brain and is primarily treated with antiepileptic drugs. In contrast, although they are likely to have a neurobiological basis, most nonepileptic seizures are understood as an involuntary, reflex-like response to distress and typically treated with psychotherapy (Brown & Reuber, 2016)..

In this report we extend this line of research by presenting the results of a phonetic analysis comparing speech samples from people with a confirmed diagnosis of epilepsy, to those with a confirmed diagnosis of NES. A recent scoping review provided by Dutta, Murray, Miller, and Grovesc (2018) looked expressly at language abilities; it seems that there is little if anything known about the possible influence of epilepsy on speech (and concomitantly, about the influence of NES on speech). Better diagnostic methods are urgently needed to reduce misdiagnoses and promote a more efficient use of investigations (Malmgren, Reuber, & Appleton, 2012). Therefore we set out to see if phonetic analysis could help to differentiate these two groups (i.e. EP vs NES). This is in some respects an extension of existing work using analysis of speech to distinguish patient groups. For example, Illes (1989) compares AD, Huntingdon's Disease and Parkinson's Disease; Croot, Ballard, Leyton, and Hodges (2012) looked at the speech of a mixture of patients with nonfluent/agrammatic PPA, and logopenic PPA. Hoffmann et al. (2010) compares patients with mild, moderate and severe AD with healthy controls, and Kato, Homma, Sakuma, and Nakamura (2015), König et al. (2015), Midi et al. (2011) and Satt et al. (2013) compare healthy controls to patients with MCI and patients with varying severity of AD. Building on extensive research into the effects of depression on speech, Cummins et al. (2015) review the different effects of depression and suicidality on speech production and the potential for acoustic analysis to distinguish between the two groups.

The acoustic analyses we present are grounded in differences in phonetic parameters known to be of relevance in the organisation of conversation such as pitch range relative to a speaker's baseline, mean intensity or loudness, and speaking rate (G. Walker, 2013). This helps ensure that the any differences identified are likely to be evident to listeners. This is an important difference from other work on patient groups which focusses on measurements which can be automated and which show statistically significant effects, regardless of the likelihood that listeners would be able to detect the differences in the speech signal. This preliminary work also addresses the gap in research into the speech functions of persons with epilepsy.

Method

This study employs conversation analytic insights with impressionistic and acoustic phonetic analysis. The data set comes from recordings of the speech of patients who had been referred for observation with video-EEG (electroencephalography) because of uncertainty about their current diagnosis. We compare 7 patients with a proven diagnosis of epilepsy, and 8 with NES. These recordings have been drawn on previously by Plug and Reuber

(2009), Plug et al. (2009a, 2009b, 2009c), and Reuber et al. (2009). As noted in Reuber et al. (144:2009), all participants were fluent English speakers, and were assessed for linguistic competence using the The Graded Naming Test and the Test for Reception of Grammar (Version 2). This study was approved by the South Sheffield Research Ethics Committee. All participants gave their informed written consent before their interview with the physician.

The recordings consist of history-taking by a neurologist (the last author), using a modified interview schedule first described in Schwabe, Reuber, Schöndienst, and Gülich (2008) and also described in Plug et al. (2009b, 2009c). The doctor conducting the history-taking interview did not know the diagnosis at the time of recording; the diagnoses were only confirmed later, after a seizure had been recorded. Our analysis focuses only on the responses to the first two questions asked by the doctor, which are “what were your expectations when you came to hospital?” and “Can you tell me about your first seizure?” The protocol for these interviews stipulated that the interviewing doctor refrain from suggesting new topics, and that he tolerate silences much longer than are normal in everyday conversation (Jefferson, 1989) as well as use continuers such as “mmhm, right” to indicate continued attention to the patient.

We have limited our current investigations to only the first two questions in part because previous literature emphasises the importance of the opening section of medical interactions (Heritage & Robinson, 2006; Robinson, 2006). In addition, Plug et al. (2009a) provide a case comparison showing that both during the opening phase of the interview and when asked about a specific seizure, Jack (EP) offers a lot of detail, whereas Barbara (NES) offers little. (All names used to refer to the data in this study are pseudonyms.) Later in the interview, when the doctor asks them about a symptom they have previously mentioned, the opposite occurs: Jack offers little detail (presumably because he has given so much earlier in the interaction) whereas Barbara offers more. This shows that people with NES need prompting to expand on their experiences whereas people with epilepsy do not. Therefore, we have proceeded on the assumption that the diagnostically relevant linguistic features of epilepsy/NES occur in the earlier part of the interaction in which patients’ contributions to the interaction are less strongly affected by the doctor’s contributions and what has happened earlier in the encounter.

Finally, the responses to the first and second questions provide a manageable amount of data: an average of 29 seconds and 50 seconds of speech, excluding pauses, for the responses to the first and second questions respectively, with a minimum combined total for responses to first and second questions of 26 seconds (average = 79 seconds). This seems to us an appropriate amount of data for preliminary examination of the two patient groups, and is supported by research on effective ways to study speakers’ pitch characteristics (which we study here). For example, Horii (1975) reports that the variability of fundamental frequency mean and median estimations decreases as sampling size (amount of speech) increases. While this decrease in variability is fast initially, it slows as sample size increases, and a sample size of approximately 14 s – much smaller than the samples we consider here – achieves an accuracy of approximately 3.2 Hz for mean fundamental frequency. Lennes, Stevanovic, Alto, and Palo (2015) also report a decrease in variability of pitch means as sampling size increases, with reduction in the standard deviation of pitch mean estimates to approximately 1 ST (semitone) by the time that 12 s of speech are considered.

Using existing transcriptions, the responses to the first and second questions from the interviewer were identified according to conversation analytic principles (Sidnell & Stivers, 2013). The ends of responses were judged by more conservative estimates than in the original studies, which considered all the patient talk up until the doctor’s next query as the ‘response’. Our analysis instead marked the end of a response when the patients’ talk veered off into other topics, eg., “but I were scared”, [Tallulah, line 11] in response to the first question about expectations, or in response to the second question about their first seizure, “think second one were just about same” [Betty, line 233], or “I’ve had encephalitis” [David, line 15]. We did this to ensure that our analysis focussed only on talk given in response to the interviewer question, partly since - as noted above - it has already been demonstrated that the content of the answers can be used to differentiate the patient groups. By constraining the talk we considered even further, we hoped to ensure that we were not analysing off-topic talk that might have a different phonetic design.

All audio recordings were normalised for loudness using *normalize* (http://normalize.nongnu.org/), bringing the volume of each recording to the same overall level. This step was taken due to differences in the clininical environments in which the recordings were made, the position of the speakers relative to the microphone and differences in recording gain. These differences can have a significant impact on the comparability of intensity measures. Each recording was adjusted by a different amount due to differences in overall loudness. However, the same adjustment was applied within each recording depending on the overall loudness of the recording, including the speech from the interviewer and from the patient. This reduces the likelihood that the normalisation procedure would mask some differences between the patient groups (cf. normalising the loudness of each patient's response in isolation, for example, or not normalising the recordings at all). Each response was annotated using Praat (Boersma & Weenink, 2018), combining visual inspection of the waveform and spectrogram with listening to the audio playback. Material was transferred manually from the transcriptions to Praat annotation files (TextGrids), marking the beginning and end of each event (utterance, or transcribed silence). Some details of the original transcription were adjusted by consensus. The TextGrids have separate tiers for the patient’s speech, the interviewer’s speech, silences which could be attributed to the patient (within-turn silences), silences which could be attributed to the interviewer, and silences which could not be attributed to either party: see Figure 1. A further tier was added later to record the number of syllables produced in each interval indicating speech produced by the patient: the contents of this tier are described below.

These annotations were used as the basis for studying the variables summarised in Table 1, for each of the 15 patients. The variables were selected for three main reasons: (1) they can be measured reliably on the available materials; (2) they are likely to be discernible by an observer (e.g. a doctor) without any special training or computer processing of audio data; (3) they give reasonable coverage of prosodic aspects of a speaker’s voice (pitch, loudness, rate of articulation).

Praat was used to prepare pitch traces of the samples, using Praat’s autocorrelation method. Pitch traces were prepared for all portions of the audio annotated as containing speech by the patient, starting and ending at the manually-placed boundary, and excluding any portions of overlapping talk by the interviewer. This step was intended to help reduce the number of pitch values which correspond to background noise, or speech by the interviewer. Pitch traces were prepared with different pitch floor and ceiling values for male and female patients, using the values suggested in the Praat manual (web version: http://www.fon.hum.uva.nl/praat/manual/Intro_4_2__Configuring_the_pitch_contour.html): a floor of 75 Hz for male patients and a ceiling of 300 Hz; a floor of 100 Hz for female patients and a ceiling of 500 Hz. Praat will only provide pitch estimates between these floor and ceiling values. All of the pitch traces were then corrected through a combination of visual and auditory inspection. Pitch values which did not reflect the patient’s speech were either moved within the constraints of Praat’s constraints on pitch editing, or removed by marking the frame(s) as unvoiced; no values were added in for frames which Praat considered to be unvoiced. Pitch values for portions where non-modal voice quality is evident and which would have given misleading representations of the speaker's normal speaking range (e.g. very low values produced during portions of creaky voice) were removed.

The corrected pitch traces were used as the basis of the intensity measures: intensity measures were taken wherever a pitch value occurred in the pitch traces. This step helps ensure that the intensity measures are only taken where voiced speech from the patient has been deemed to occur, and reduces the likelihood that the intensity measures reflect only background noise.

The extent of labelled intervals on the patient speech tier in the annotations described above were used as the basis for the speechTime variable, starting and ending at the boundary once moved to the nearest zero-crossing. The variable speechSylls was determined by totalling up the number of syllables produced by the speaker in all labelled intervals on the patient speech tier. Syllables were counted on the basis of what could be heard, rather than a canonical production, for example "to the door" = 3 syllables, but produced as "tut door" = 2. Any queries were resolved through discussion.

The variables described above all relate aspects of the patient's speech in their responses to the first and second question from the interviewer. There is another relevant set of variables to analyse, however; those of 'non-speech', ie., pauses in the patient's responses. Pauses are clearly an audible feature of the responses, and pauses have been shown to be important in discriminating between other patient groups. These variables are summarised in Table 2.

In the remainder of the paper we present a descriptive statistical analysis of these variables; namely, the minimum, maximum, mean, median and standard deviation for each variable for each of the groups, as well as plots showing the results for individual speakers from each group. Skewness in the distribution of pitch and intensity values for each speaker is described in the Discussion section. As already noted, the aim of this particular research project is to investigate audible differences in the pitch, loudness, and/or timing of speech produced by people with epilepsy compared to people with non-epileptic seizures; therefore we leave for other researchers any more complex mathematical comparisons between the measured values. Additionally, we must point out that our pool of secondary data was small: only 15 speakers in total, 7 EP and 8 NES. Given these small numbers, we err on the side of caution and make no claims about statistical significance.

Results

The results reported below are based on an analysis of combined responses to the first and the second question from the doctor. Table 3 presents the minimum, maximum, mean, median and standard deviation for each variable for each of the groups.

Visual representations of variables

Plots for each variable listed in Table 1 are shown in Figure 2. In each, the lowest values for the given variable are on the left of the x-axis, and the values increase moving to the right.

It is possible to add together the rankings on single variables to give an overall ranking for each patient. Figure 3 shows the output of combining the rankings on each of the variables in Figure 1, except speechTime. This has been excluded as this variable is effectively the same as speechSylls (*r* = 0.99), and the rankings almost identical (the only change is that the patients ranked as 7, 8 and 9 on speechSylls are ranked as 9, 7, 8 on speechTime: see Figure 2d and 2e). On each variable, the highest ranked speaker on that variable was given a score of 15; the lowest ranked speaker was given a score of 1. The higher the score on each variable, the 'bigger' the speech (biggest pitch range, highest median relative to the speaker's baseline, loudest, most syllables, fastest). Note that the distance between adjacent rankings on a variable in terms of ST, dB, sylls or sylls/s, is factored out when totalling up rankings. In other words, the results in Figure 3 are based on the speech of the patients, but the figures are not a direct measure of speech. Instead, this is a method for comparing two groups, and for comparing individuals within those two groups.

Pitch and intensity profiles

In the plots shown in this section pitch values are in 1 ST bins; intensity values are in 1 dB bins. The plots show all values in the results; the x-axis ranges in the plots are the same for all participants.

Plots by speaker

The plots in Figure 4-7 show pitch and intensity profiles for each speaker, arranged by diagnosis (EP vs. NES).

Plots by patient group

Figure 8 and 9 bring together pitch and intensity measures by all individuals in each patient group. Note that there is no normalisation for the number of measures for each speaker: in other words, a speaker who generates more measures will have a bigger visual impact on these plots than a speaker generating fewer.

Pausing

Figure 10 shows plots of the variables related to pausing described in Table 2.

Discussion

Based on the results in the previous section, it is our contention that few if any straightforward audible distinctions between EP and NES are immediately apparent. This may be explained by the fact that there are subgroups of patients with NES which can be separated on the basis of their underlying psychopathology, as suggested by e.g., Brown et al. (2013); Reuber et al. (2003). Similarly, patients with EP are not a homogeneous group in terms of certain clinical characteristics (such as neurological impairment, the presence of psychiatric comorbidity, or the location in the brain where the seizure originates).Our small sample is not controlled for these variables. In what follows we first discuss some of the findings that may be of potential diagnostic relevance, and then discuss our findings/lack thereof from a linguistic standpoint. We do not cover each of the eight variables already described in the methodology and the results section, but instead confine our discussion to those the show some (admittedly slight) difference between EP and NES.

Pitch range and amount of speech produced

Figure 2a shows that, in our dataset, the narrowest pitch ranges are used by speakers with NES (see TL and TM). Additionally, Figure 3 shows that those who come out lowest on combined rankings on all parameters also have a diagnosis of NES (TM, PA, TL). Finally, for amount of speech, measured either in raw time or in syllables produced, patients with NES cluster at both ends of the spectrum; in other words, they produce both the least and the most speech (see Figure 2d, and Figure 2e). Based on these findings, future research might focus on whether patients with NES exhibit an overall 'depression' in the phonetic production of their speech compared to patients with EP. The findings regarding the amount of speech produced might indicate that there is a 'typical' amount of time that a patient with EP takes to respond to these questions, and that outliers from that indicate a non-EP diagnosis. Both of these areas could repay additional research, as they would be fairly easy for anyone (not just trained phoneticians) to hear and recognise.

It is conceivable that the two different phonetic patterns for people with NES reflect the well-recognised clinical heterogeneity within the NES group: some but not all people with NES will have clinical features of conditions such as anxiety, depression or post-traumatic stress disorder.

Pitch and intensity profiles

Numerical measures of the skewness in the distribution of pitch and intensity values for each speaker, arranged by group (EP vs. NES) are shown in Table 4.

The distribution of pitch values seems to be more skew for speakers in the EP group than for speakers in the NES group. The mean skewness for speakers in the EP group = 0.92, while the mean skewness for speakers in the NES group = 0.58. To put these measures in context, Bulmer (1979) suggests as a rough guide that if skewness is greater than 1 then the distribution is highly skew; if skewness is between 0.5 and 1 then the distribution is moderately skew; if skewness is between 0 and 0.5, then the distribution is fairly symmetrical.

When pitch values from all speakers in each group are combined, as in Figure 8, some difference in skew in the distribution of pitch values is evident between the two groups. While the distribution of pitch values is skew to the right for both groups, the distribution of pitch values is more skew to the right for the EP group than for the NES group (0.8 for the EP group, 0.67 for the NES group).

In terms of how the speakers in the two groups may sound, these general tendencies in the distribution of pitch values suggest at least two things. First, that speakers in both groups tend to produce their speech towards the bottom of their pitch range. This is consistent with the Lennes et al. (2015): working with samples of speech from 40 individuals (20 female, 20 male), they report that the distributions of pitch values tend to be skew to the right. Second, these general tendencies suggest that speakers in the EP group produce more of their speech towards the top of their pitch range than speakers in the NES group.

With one exception, the distribution of intensity values for all speakers in both groups is skew to the right. The difference in mean skewness between the two groups is very small (EP = 0.36, NES = 0.44), and this is reflected in the similarity between Figure 9a and 9b. Whether these differences in the distribution of pitch and intensity values are reflected in listeners' perceptions, and whether these patterns are evident in speech beyond these samples, is another matter. We are wary of putting too much stock in these small statistical differences, especially where their relationship with what can be heard is unclear (Foulkes, Docherty, Shattuck-Hufnagel, & Hughes, 2018).

Pausing

There is no obvious distinction to be made between EP and NES on the basis of these three variables. However, we note the following: (i) the speakers who pause both the fewest (TM, CH, TL) and the most times (SU, BA, LA) are in the NES group (Figure 10a); (ii) the speakers with the shortest average pause durations are from the EP group (SM, ZA, Figure 10b); (iii) the speakers with the lowest pause-to-speech ration are from the EP group (SM, ZA; Figure 10c).

Conclusions, and further work

The findings we have reported in this paper are somewhat inconclusive, and in places point to a negative finding, i.e., no obvious straightforward relationship between a phonetic parameter and a diagnostic group (which can of course provide a valuable addition to knowledge). Some possible explanations for these findings are: (i) there are no phonetic differences between these two patient groups; (ii) there are differences between these two patient groups, but these are the wrong variables to capture those differences; (iii) there are differences between these two patient groups, these are the right variables to capture those differences, but the measurements are unreliable (e.g. sample size too small, errors in measuring).

With regard to the relatively small sample size analysed here, we note that the preparation of audio data collected in the field is a time-consuming process. However, the high ecological validity of such data makes the process worthwhile. Because of the time involved, it seems appropriate to conduct a pilot study in a modest but technically manageable patient group to establish whether this method is worth pursuing in a larger sample. Given the lack of signal from this study, we come to the conclusion that phonetic analysis that does not take account of the context is unlikely to generate findings of differential diagnostic value.

We would in fact argue that contrary to what may be lay expectation, there is not much linguistically-motivated evidence to support the idea that one could diagnose a particular illness on the basis of acoustic material alone. The work done so far on the differential diagnosis of epilepsy in fact supports the opposite, showing that the lexico-semantic-pragmatic content of talk – eg., the use of metaphor, (Plug et al., 2009b); the use of the term 'seizure' (Plug et al., 2009c) – is where the relevant differences lie.

Phonetic parameters such as pitch and intensity are not produced independently of the words that carry them; indeed we have argued elsewhere against the idea that any phonetic constellations have invariant meanings that can be simply overlaid on the lexis (Curl, Local, & Walker, 2006; Ogden & Walker, 2013; G. Walker, 2013; T. Walker, 2014). Rather, Benjamin and Walker (2013), Curl (2005), Curl et al. (2006), T. Walker and Benjamin (2017) show how phonetic patterns combine with what is said to differentiate actions; therefore it may be the case that if analysed in tandem with the content of the talk, the phonetic patterns may contribute to the differentiation of these or other patient groups, although we have found little evidence that the phonetic productions on their own could do so.

As might be anticipated, this preliminary investigation has some limitations. As noted previously, both EP and NES are heterogenous groups. Our small sample might contain people with a variety of different underlying and/or co-occuring conditions.

In terms of what talk we analysed, we only investigated the combined responses to the first and second questions asked during the interview. Additionally, we limited the response only to talk that directly related to the question. We could expand the amount of speech considered to all talk up to the interviewer's next question; additionally, we could also extend the analysis to all responses to all questions. However, the grounds on which we selected the subset of responses were well-motivated, as these responses have been shown previously to contain the kind of talk that that differentiates the patient groups.

Another limitation, not easily overcome in the short term, is the variability in question responses that we see in this dataset. Although this data contains responses that differentiated the diagnostic groups of EP and NES, we did note that the interviewer used a variety of grammatical forms when posing the question about 'expectations' as well as asking for details of the first seizure. We have shown elsewhere that using a computer to ask invariant questions generates responses that differentiate people with neurodegenerative disease from those with benign memory complaints just as well as questioning from a trained neurologist (T. Walker et al., 2018). However, the responses are in some ways more linguistically similar both to each other and within groups, and thus more amenable to comparison. Although the aim of our research is not the computerised acoustic analysis of speech, having the questions posed in an invariant yet still acceptable way, as might be done via a computer interface, might provide a more constrained dataset.

Additionally, we could look in more fine-grained detail at the materials we have collected. For example, we could consider the distribution of values comprising each variable over the time-course of the responses e.g. pitch movements (rises, falls), changes in articulation rate (speeding up, slowing down). What we have provided here only shows the combined responses to first and second questions; we could look at first question response vs. second question response vs. combined question responses, to see what differences there may be. Extending the analysis in this way would address our concerns about how the phonetic parameters we are investigating are implicated in other conversational organisation work, eg., turn-taking and aspects of prosodic structure, as such work may be disproportionately represented in the samples. That is to say, the length and also number of turns taken by different speakers varies greatly, and some of the phonetic parameters we are looking at to differentiate the diagnostic groups play a role in the regulation of turn-taking.

Finally, it might be argued that we should conduct more sophisticated statistical analyses; however, we are wary of doing this, especially if the relationship between the statistical analysis and what can be heard becomes unclear (Foulkes et al., 2018). Going forward, we instead hope to relate this to work on phonetic correlates of depression given that HADS (Hospital Anxiety and Depression Scale; Zigmond & Snaith, 1983) scores and other background data is available for most of these speakers. This seems especially worthwhile given the comorbidity of depression within both diagnostic groups.

References

Benjamin, T., & Walker, T. (2013). Managing problems of acceptability through high rise-fall repetitions. *Discourse Processes*, 50 (2), 107-138. doi:10.1080/0163853x.2012.739143

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/.

Brown, R. J., & Reuber, M. (2016). Towards an integrative theory of psychogenic non-epileptic seizures (PNES). *Clinical Psychology Review*, 47, 55-70.

Brown, R. J., Bouska, J. F., Frow, A., Kirkby, A., Baker, G. A., Kemp, S., et al. (2013). Emotional dysregulation, alexithymia, and attachment in psychogenic nonepileptic seizures. *Epilepsy and Behaviour,* 29 (1),178-83.

Bulmer, M. (1979). *Principles of statistics*. Mineola, NY: Dover Publications, Inc.

Croot, K., Ballard, K., Leyton, C. E., & Hodges, J. R. (2012). Apraxia of speech and phonological errors in the diagnosis of nonfluent/agrammatic and logopenic variants of primary progressive aphasia. *Journal of Speech Language and Hearing Research*, 55 (5), S1562-S1572. doi:10.1044/1092-4388(2012/11-0323)

Cummins, N., Scherer, S., Krajewski, J., Schnieder, S., Epps, J., & Quatieri, T. F. (2015). A review of depression and suicide risk assessment using speech analysis. *Speech Communication*, 71, 10-49. doi:10.1016/j.specom.2015.03.004

Curl, T. S. (2005). Practices in other-initiated repair resolution: The phonetic differentiation of ‘repetitions’. *Discourse Processes*, 35 (1), 1-43. doi:10.1207/s15326950dp3901 1

Curl, T. S., Local, J., & Walker, G. (2006). Repetition and the prosody-pragmatics interface. *Journal of Pragmatics*, 38 (10), 1721-1751. doi:10.1016/j.pragma.2006.02.008

Dutta, M., Murray, L., Miller, W., & Grovesc, D. (2018). Effects of epilepsy on language functions: Scoping review and data mining findings. *American Journal of Speech-Language Patholog*y, 27, 350-378. doi:10.1044/2017 AJSLP-16-0195

Foulkes, P., Docherty, G., Shattuck-Hufnagel, S., & Hughes, V. (2018). Three steps forward for predictability. Consideration of methodological robustness, indexical and prosodic factors, and replication in the laboratory. *Linguistics Vanguard*, 4 (s2). doi:10.1515/lingvan-2017-0032

Heritage, J., & Robinson, J. D. (2006). The structure of patients’ presenting concerns: Physician’s opening questions. *Health Communication*, 19 (2), 89-102. doi:10.1207/s15327027hc1902 1

Hoffmann, I., Nemeth, D., Dye, C. D., Pakaski, M., Irinyi, T., & Kalman, J. (2010). Temporal parameters of spontaneous speech in Alzheimer’s disease. *International Journal of Speech-Language Pathology*, 12 (1), 29-34. doi:10.3109/17549500903137256

Horii, Y. (1975). Some statistical characteristics of voice fundamental frequency. *Journal of Speech, Language, and Hearing Research*, 18 (1), 192-201. doi:10.1121/1.1981923

Illes, J. (1989). Neurolinguistic features of spontaneous language production dissociate 3 forms of neurodegenerative disease - Alzheimer’s, Huntington’s, and Parkinsons. *Brain and Language*, 37 (4), 628-642. doi:10.1016/0093-934X(89)90116-8

Jefferson, G. (1989). Preliminary notes on a possible metric which provides for a ‘standard maximum’ silence of approximately one second in conversation. In D. Roger & P. Bull (Eds.), *Conversation: An interdisciplinary perspective* (pp. 166-196). Clevedon: Multilingual Matters.

Jenkins, L., & Reuber, M. (2014). A Conversation Analytic Intervention to Help Neurologists Identify Diagnostically Relevant Linguistic Features in Seizure Patients' Talk. *Research on Language and Social Interaction*, 47 (3), 266-79.

Jenkins, L., Cosgrove, J., Chappell, P., Kheder, A., Sokhi, D., & Reuber, M. (2016). Neurologists can identify diagnostic linguistic features during routine seizure clinic interactions: results of a one-day teaching intervention. *Epilepsy and Behaviour*, 64 (Pt A), 257-61.

Kato, S., Homma, A., Sakuma, T., & Nakamura, M. (2015). Detection of mild Alzheimer's Disease and mild cognitive impairment from elderly speech: Binary discrimination using logistic regression. In *Proceedings of the 37th annual international conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5569-72). doi:10.1109/EMBC.2015.7319654

König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P. H., & David, R. (2015). Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1 (1), 112-124. doi:10.1016/j.dadm.2014.11.012

Kotsopoulos, I.A., de Krom, M.C., Kessels, F.G., et al. (2003). The diagnosis of epileptic and non-epileptic seizures. *Epilepsy Research*, 57:59–67.

Lennes, M., Stevanovic, M., Alto, D., & Palo, P. (2015). Comparing pitch distributions using Praat and R. *The Phonetician*, 111-112, 35-53.

Malmgren, K., Reuber, M., & Appleton, R. (2012). Differential diagnosis of epilepsy. In S. Shorvon, R. Guerrini, M. Cook, & S. Lhatoo (Eds.), *Oxford textbook of epilepsy and epileptic seizures* (pp. 81-94). doi:10.1093/med/9780199659043.003.0008

Midi, I., Dogan, M., Pata, Y. S., Kocak, I., Mollahasanoglu, A., & Tuncer, N. (2011). The effects of verbal reaction time in Alzheimer's disease. *Laryngoscope*, 121 (7), 1495-1503. doi:10.1002/lary.21753

Ogden, R., & Walker, T. (2013). Phonetic resources in the construction of social actions. In B. Szczepek Reed & G. Raymond (Eds.), *Units of talk – units of action* (pp. 277-312). Amsterdam: John Benjamins.

Plug, L., & Reuber, M. (2009). Making the diagnosis in patients with blackouts: It's all in the history. *Practical Neurology*, 9 (1), 4-15. doi:10.1136/jnnp.2008.161984

Plug, L., Sharrack, B., & Reuber, M. (2009a). Conversation analysis can help in the distinction of epileptic and non-epileptic seizure disorders: A case comparison. *Seizure*, 18, 43-50. doi:10.1016/j.seizure.2008.06.002

Plug, L., Sharrack, B., & Reuber, M. (2009b). Seizure metaphors differ in patients' accounts of epileptic and psychogenic non-epileptic seizures. *Epilepsia*, 50 (5), 994-1000. doi:10.1111/j.1528-1167.2008.01798.x

Plug, L., Sharrack, B., & Reuber, M. (2009c). Seizure, fit or attack? The use of diagnostic labels by patients with epileptic and non-epileptic seizures. *Applied Linguistics*, 31 (1), 94-114. doi:doi.org/10.1093/applin/amp012

Reuber, M., Monzoni, C., Sharrack, B., & Plug, L. (2009). Using interactional and linguistic analysis to distinguish between epileptic and psychogenic non-epileptic seizures: A prospective, blinded multirater study. *Epilepsy and Behavior*, 16 (1), 139-144. doi:10.1016/j.yebeh.2009.07.018

Reuber, M., Fernández, G., Bauer, J., Helmstaedter, C., & Elger, C.E. Diagnostic delay in psychogenic nonepileptic seizures. *Neurology,* 58 (3), 493-495.

Reuber, M., Pukrop, R., Derfuss, R., Bauer, J., & Elger, C. E. (2003). Multidimensional assessment of personality in patients with psychogenic nonepileptic seizures. *Journal of Neurology, Neurosurgery & Psychiatry*, 75, 743-48.

Robinson, J. D. (2006). Soliciting patients' presenting concerns. In J. Heritage & D. W. Maynard (Eds.), *Communication in medical care: Interactions between primary care physicians and patients* (pp. 23-47). doi:10.1017/cbo9780511607172.004

Satt, A., Sorin, A., Toledo-Ronen, O., Barkan, O., Kompatsiaris, I., Kokonozi, A., & Tsolaki, M. (2013). Evaluation of speech-based protocol for detection of early-stage dementia. In F. Bimbot, C. Cerisara, C. Fougeron, G. Gravier, L. Lamel, F. Pellegrino, & P. Perrier (Eds.), *Proceedings of the 14th annual conference of the International Speech Communication Association (Interspeech 2013)* (pp. 1692-1696). Lyon, France.

Schwabe, M., Reuber, M., Schöndienst, M., & Gülich, E. (2008). Listening to people with seizures: How can Conversation Analysis help in the differential diagnosis of seizure disorders. *Communication and Medicine*, 5 (1), 59-72. doi:doi.org/10.1558/cam.v5i1.59

Sidnell, J., & Stivers, T. (Eds.). (2013). *The handbook of conversation analysis*. Oxford: Wiley-Blackwell.

Walker, G. (2013). Phonetics and prosody in conversation. In J. Sidnell & T. Stivers (Eds.), *The handbook of conversation analysis* (pp. 455-474). doi:10.1002/9781118325001.ch22

Walker, T. (2014). Form □ function: The independence of prosody and action. *Research on Language and Social Interaction*, 47 (1), 1-16. doi:10.1080/08351813.2014.871792

Walker, T., & Benjamin, T. (2017). Phonetic and sequential differences of other-repetitions in repair initiation. *Research on Language and Social Interaction*, 50 (4), 330-347. doi:10.1080/08351813.2017.1340717

Walker, T., Christensen, H., Mirheidari, B., Swainston, T., Rutten, C., Mayer, I., Blackburn, D., & Reuber, M. (2018). Developing an intelligent virtual agent to stratify people with cognitive complaints: A comparison of human–patient and intelligent virtual agent–patient interaction. *Dementia*. doi:10.1177/1471301218795238

Zigmond, A., & Snaith, R. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67 (6), 361-370. doi:10.1111/j.1600-0447.1983.tb09716.x

Tables

name	axis label	description	unit
pRangeST	Pitch range (ST)	difference between lowest pitch value (baseline) and the highest (topline)	semitones
pMedianST	Pitch median (ST re. baseline)	difference between the baseline and the median pitch value	semitones
iMean	Mean intensity (dB)	mean of all intensity measures for all voiced frames in pitch traces	deciBels
speechTime	Speech (s)	duration of all audible vocalisation, including inbreaths and restarts, but excluding any silent portions	seconds
articRate	Articulation rate (sylls/s)	speechSylls divided by speechTime	syllables per second

Table 1: Summary of variables

name	axis label	description
numberOfPauses	Number of pauses	number of pauses in the patient's responses
averagePauseTime	Average pause duration (s)	duration of all pauses divided by numberOfPauses, in seconds
pauseToSpeech	Pause to speech	duration of all pauses divided by speechTime; how much of the response is pause as opposed to speech; 1 = equal amounts of pause and speech; <1 = less pause than speech; >1 = more pause than speech

Table 2: Summary of variables relating to pausing

Variable name	minimum		maximum		mean		median		SD	
	EP	**NES**	**EP**	**NES**	**EP**	**NES**	**EP**	**NES**	**EP**	**NES**
pRange (ST)	9.6	5.3	22.5	14.8	13.2	11.2	11.6	12.3	4.4	3.3
pMedian (ST)	2.5	2.0	7.9	7.1	4.6	5.0	4.6	5.6	2.0	1.9
iMean (dB)	62.4	62.3	72.3	70.8	68.0	66.2	68.0	67.0	3.0	3.2
speechTime (sec)	40.0	23.9	141.2	178.4	76.3	80.9	63.5	59.6	36.2	61.2
articRate (syll/sec)	4.0	3.9	5.4	5.3	4.5	4.6	4.4	4.6	0.5	0.5
numberOfPauses	24	13	70	91	37.7	43.9	35	33	15.2	30.9
avgPauseTime (sec)	0.52	0.62	1.10	1.32	0.82	0.82	0.87	0.75	0.21	0.24
pauseToSpeech ratio	0.09	0.26	0.87	0.96	0.51	0.48	0.49	0.42	0.29	0.21

Total number of EP speakers: 7; total number of NES speakers: 8

Table 3: Numeric values for pitch, intensity, speech timing, pauses

group	gender	name	pitch	intensity
EP	female	Samantha	0.79	0.32
		Sandra	0.22	0.24
	male	Carl	0.2	0.24
		David	1.52	0.58
		Jack	1.27	0.64
		Ken	1.34	0.29
		Zack	1.11	0.19
		mean	**0.92**	**0.36**
NES	female	Barbara	0.71	0.32
		Betty	-0.12	0.28
		Laura	0.69	0.23
		Patsy	0.7	0.98
		Sue	0.84	0.16
		Tallulah	-0.03	-0.08
		Tammy	1.3	1.06
	male	Chris	0.51	0.54
		mean	**0.58**	**0.44**

Table 4: Skewness of the distribution of pitch and intensity values, arranged by patient group and by gender

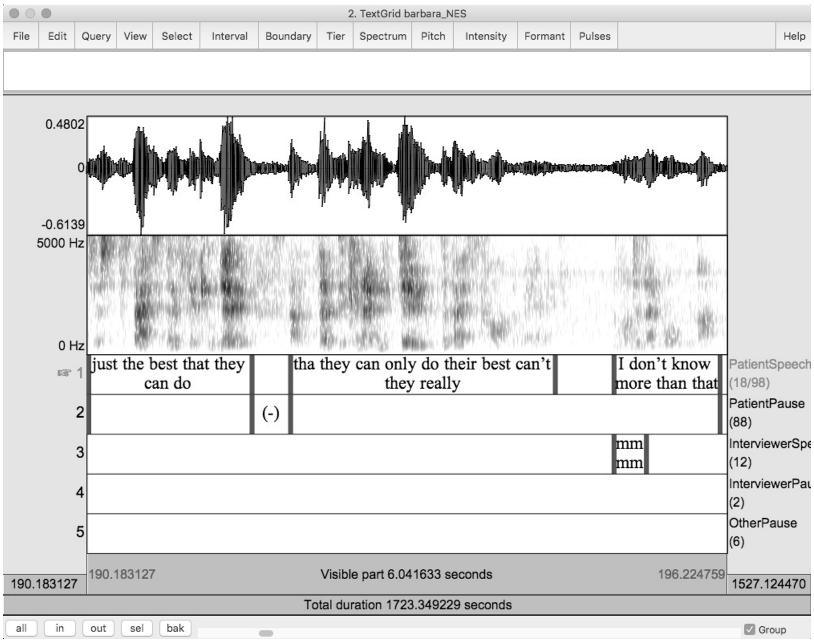


Figure 1: Annotations of files using Praat

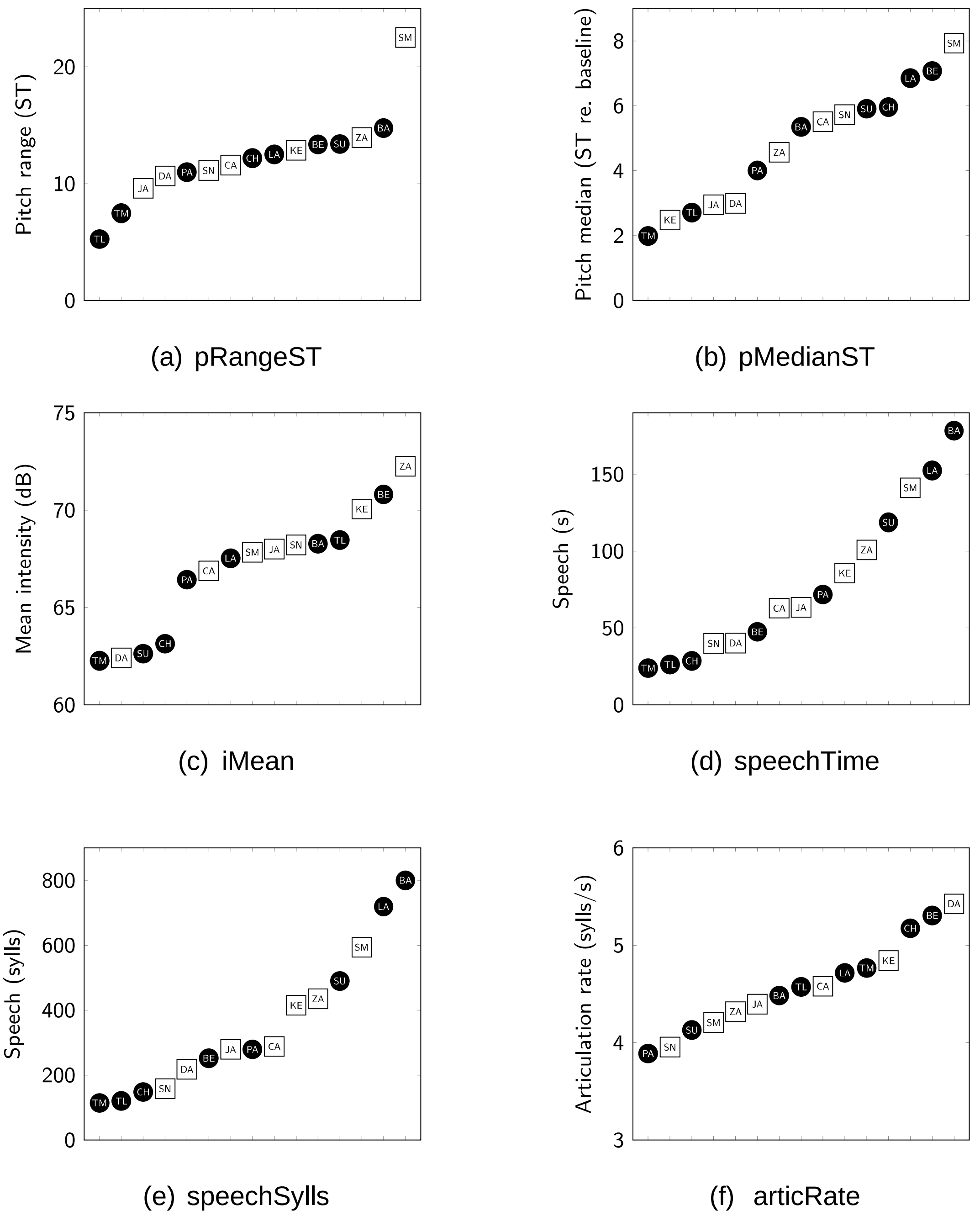


Figure 2: Single variables, AB = EP AB = NES, letters identify speakers

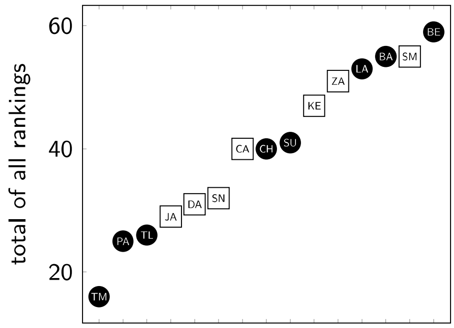


Figure 3: Total of rankings on each variable, [AB] = EP (AB) = NES, letters identify speakers

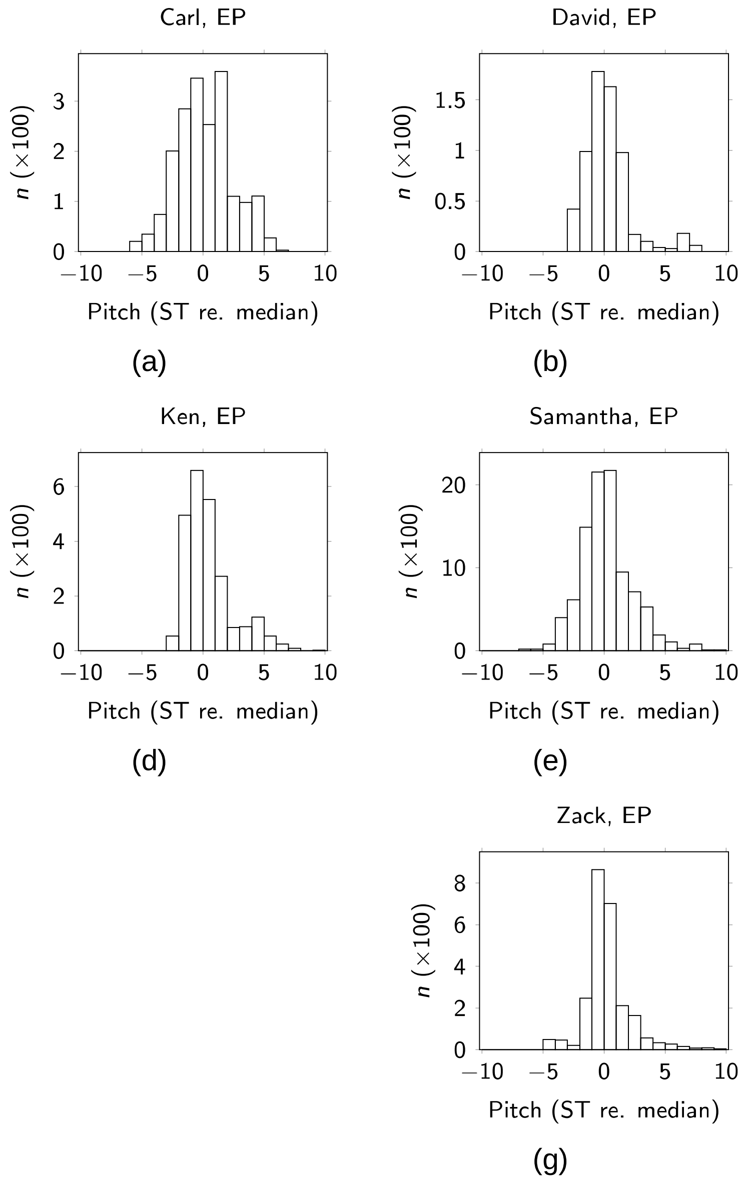


Figure 4: Pitch profiles, confirmed diagnosis of epilepsy

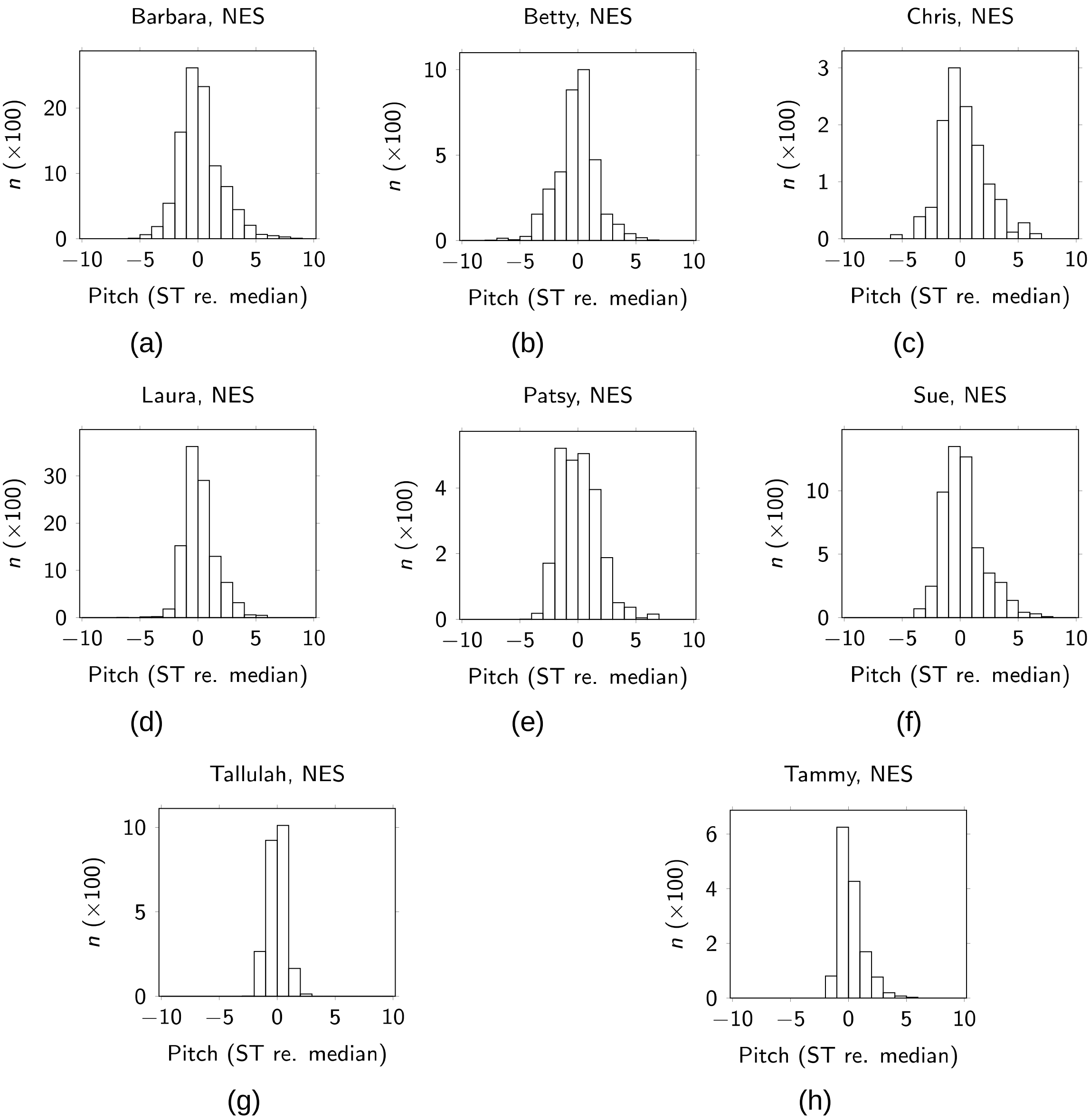


Figure 5: Pitch profiles, confirmed diagnosis of non-epileptic seizures

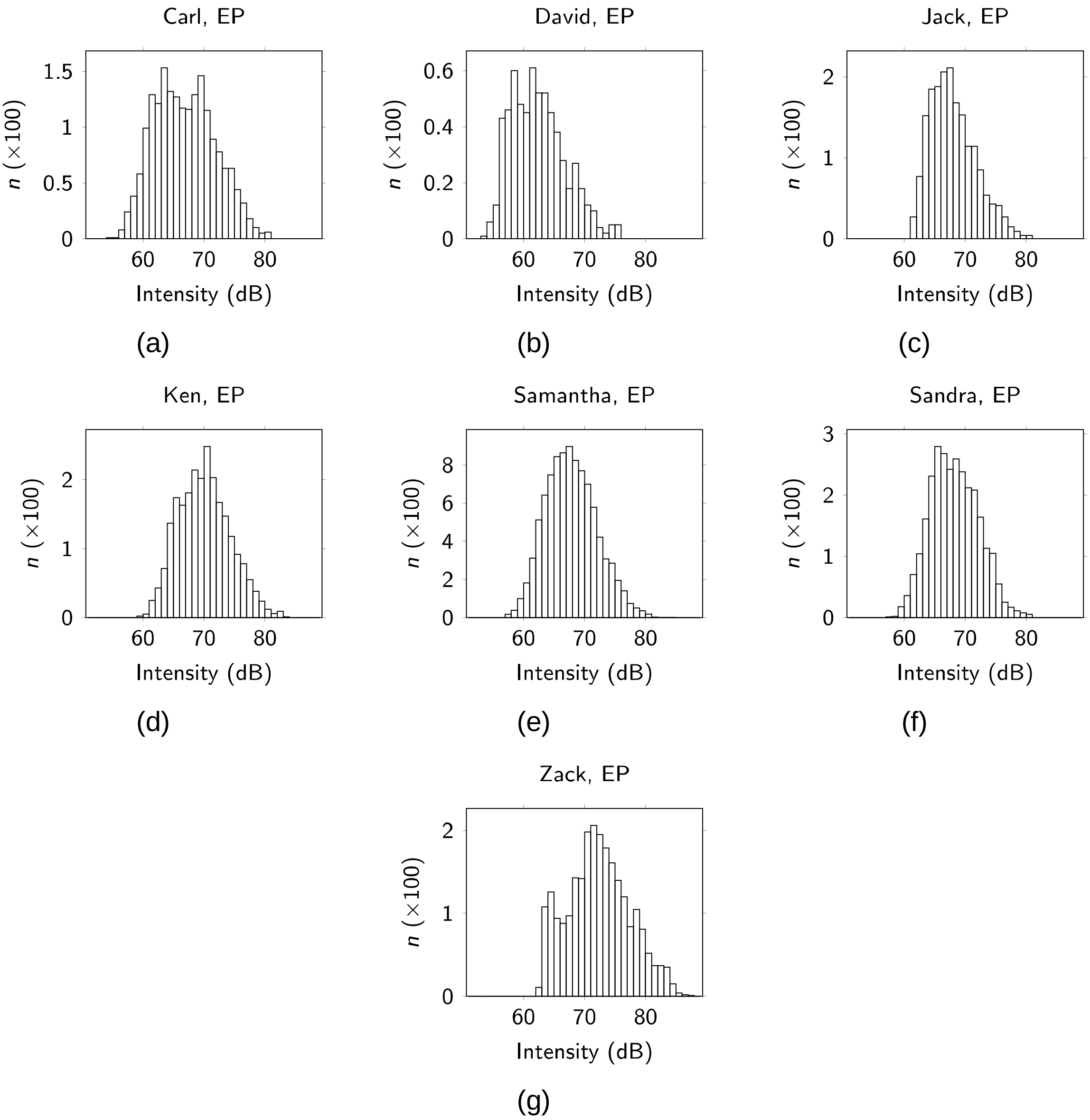


Figure 6: Intensity profiles, confirmed diagnosis of epilepsy

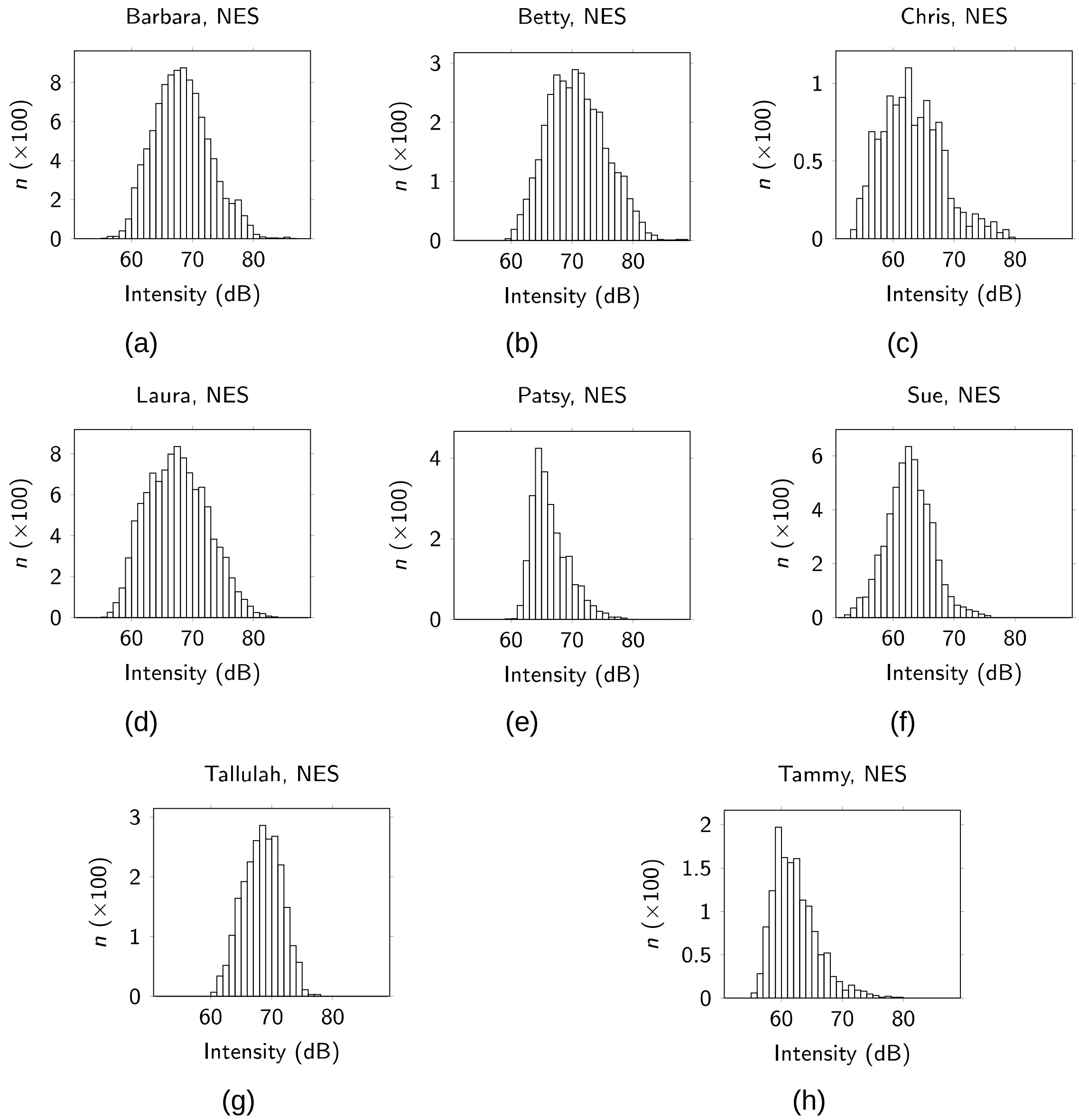
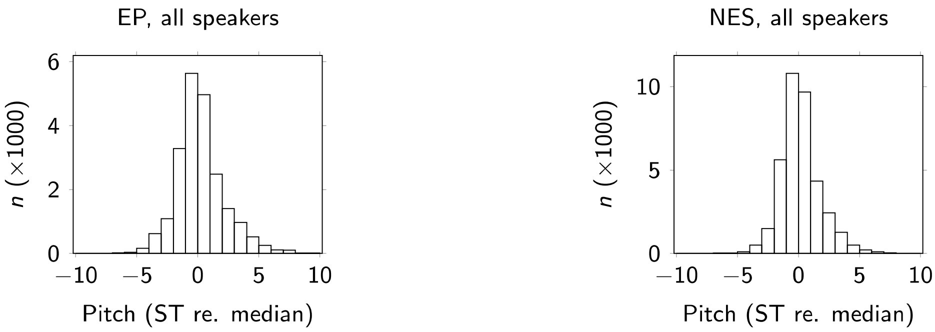
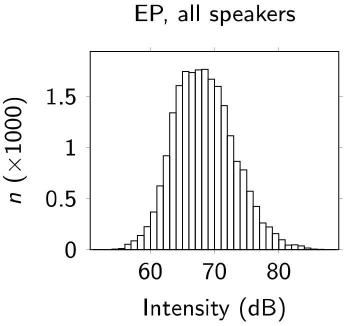


Figure 7: Intensity profiles, confirmed diagnosis of non-epileptic seizures

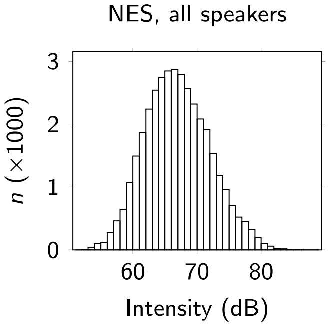


(a) (b)

Figure 8: Pitch values, all speakers

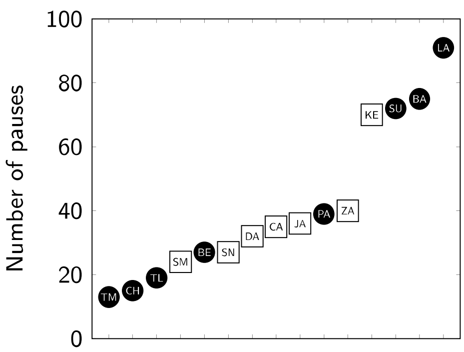


(a)

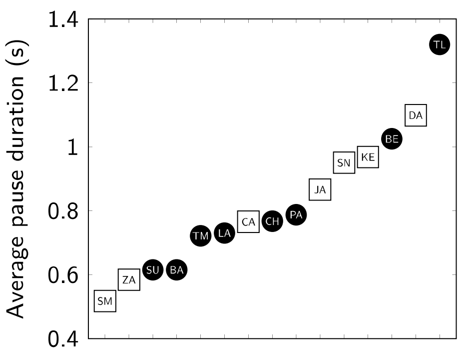


(b)

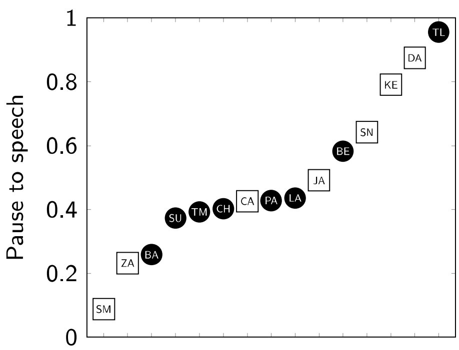
Figure 9: Intensity values, all speakers



(a) numberOfPauses



(b) averagePauseTime



(c) pauseToSpeech

Figure 10: Variables relating to pausing, AB = EP AB = NES, letters identify speakers