Classifying Acoustic Signals into Phoneme Categories: Average and Dyslexic Readers Make Use Of Complex Dynamical Patterns

Fred Hasselman[1,2]
[1]School of Pedagogical and Educational Science, Radboud University Nijmegen
[2]Behavioural Science Institute: Learning and Plasticity, Radboud University Nijmegen

*Corresponding Author:

Fred Hasselman
School of Pedagogical and Educational Science
Montessorilaan 3
P.O. Box 9104

6500 HE Nijmegen
The Netherlands
f.hasselman@pwo.ru.nl

Key-words:
Developmental Dyslexia, Speech Perception, Recurrence Quantification Analysis, Rise Time Perception Deficit, Auditory Temporal Processing Deficit, Pattern Classification, Strong Inference, Theory evaluation

ACKNOWLEDGEMENTS:

I would like to thank Anna Bosman and Ludo Verhoeven for their comments and feedback on a previous version of this manuscript.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Abstract

Aetiologies of developmental dyslexia often assume a deficit in auditory processing may be causally entailed in the specific learning disorder. The purpose of this study is to compare a number of assumed auditory features that are supposed to evidence the account given by such aetiologies under conditions of strong inference. To do so, the relevant acoustic features were extracted from the same set of artificial speech stimuli that lie on a /bAk/-/dAk/ continuum. Features were tested on their ability to enable a simple classifier (Quadratic Discriminant Analysis) to reproduce the observed classification performance of average and dyslexic readers in a speech perception experiment. The 'classical' features examined were based on component process accounts of developmental dyslexia such as the supposed deficit in Envelope Rise Time detection and the deficit in the detection of rapid changes in the distribution of energy in the frequency spectrum (formant transitions). Studies examining these temporal processing deficit hypotheses remarkably do not employ measures that quantify the temporal dynamics of stimuli. It is shown that measures based on quantification of the dynamics of complex, interaction-dominant systems enable QDA to classify the stimuli almost identically as dyslexic and average reading participants. It seems unlikely that participants used any of the features that are traditionally associated with accounts of (impaired) speech perception that assume classifying speech stimuli amounts to a linear additive interaction of component processes that each parse the acoustic signal independent of one another.

1. It's about Time! Or is It?

Many aetiologies of developmental dyslexia assume some deficit in auditory processing may be causally entailed in the difficulty with acquiring proficient levels of reading and spelling ability experienced by a small percentage of the population (e.g., Ramus, 2004). The nature of the features of the acoustic signal that are assumed to be able to evidence such deficient components (e.g., phoneme representations, allophones) or component processes (e.g., frequency sweep detection, rise time perception) varies greatly between aetiologies. The purpose of this study is to compare a number of such features under conditions of *strong inference*. The goal is to examine whether average and dyslexic readers actually use *these* particular features to arrive at a particular classification of a speech stimulus. Three types of measures will be examined that represent different distinguishing features of the speech signal, however, we will not construct stimuli that exclusively represent these measures as is common in auditory and speech perception studies (see e.g. Boets, Ghesquière, van Wieringen, & Wouters, 2007; Pasquini, Corriveau, & Goswami, 2007). Instead, we will extract all measures from one and the same set of stimuli and analyse which measures yield a classification response (by a simple classifier algorithm) that is most similar to that observed in the participants.

The measures used in this study can be extracted from any continuous signal (sampled, synthesised, or generated otherwise), but are very different in the type of information they are thought to capture, or more suitably: represent. The first are *Component Process Measures*, derived from the signal because of their supposed importance in

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

contemporary theoretical assumptions about deficient components of cognitive or sensorimotor processes related to developmental dyslexia and speech perception. They represent the Component Dominant family of dyslexia ontology. The second type of measure are *Periodicity Measures,* derived from (linear) transforms or decompositions of the signal, usually used in other contexts to express the average periodicity, harmonicity or regularity of the 'true' signal (see e.g. Guiard, 1993 for an application to harmonic movements). These measures quantify such periodic changes of the variable in question over time. The third are *Interaction Dominant Dynamics Measures* derived from nonlinear time series analyses that have a wide range of applications in the general study of complex systems, but have recently successfully been used to detect abnormal speech due to pathology or disease (cf. Little, McSharry, Roberts, Costello, & Moroz, 2007). These measures are hypothesised to capture information about the nonlinear and turbulent airflows generated by complex gestures of the human speech apparatus (cf. Little et al, 2007). They may be considered high-level, or macro-level measures that try to quantify the dynamics of the entire speech signal, not just one component. The latter two types of measure have not been the focus of studies on dyslexia and speech perception. This is quite remarkable since these measures are tailor made to test claims of deficits in detecting complex dynamic frequency or amplitude patterns present in the speech signal. The more common *Component Process Measures* quantify change over time in a peculiar way, more like a nominal variable that can be on or off in a stimulus (F2 rate of frequency change is high or low; Rate of change of envelope modulation is high or low). This is not the same as quantifying the dynamics of a continuous signal.

What follows will be an introduction to these measures and an analysis of their ability to actually serve as features to classify speech stimuli in a similar fashion as observed in the performance of average and dyslexic readers in simple labelling experiments of those stimuli.

2. Component Process Measures: What does Temporal refer to?

The "temporal" auditory processing deficit hypotheses concern properties or information in auditory stimuli that cannot, due to the rate with which the information changes over time, be properly perceived by the person afflicted with the deficit. There are two major deficit hypotheses of this kind: The auditory temporal processing deficit hypothesis (ATPDH; Farmer & Klein, 1995; Tallal, 2004) and the rise time perception deficit hypothesis (RTPDH) proposed by Goswami and colleagues (see e.g. Goswami, Fosker, Huss, Mead, & Szűcs, 2010; Goswami et al., 2002).

The ATPDH states that speech stimuli with rapid transient spectral elements are processed less accurately because such elements occur too *fast* to be perceived by people with the processing impairment. In fact, the claim is not limited to spectral features, but pertains to any sequence of auditory stimuli presented in rapid succession. Tests that have been employed to reveal this deficit are for instance temporal order judgements (e.g. Pasquini et al., 2007) and auditory gap (or threshold) detection (Boets et al., 2007; Corriveau,

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Pasquini, & Goswami, 2007). There is also evidence from neuroscience that seems to point to anomalous functional responses to rapid auditory stimuli (Temple et al., 2000) or an "asynchrony" in the speed of processing between auditory and visual modalities (Breznitz, 2003). Note that essentially, these are two different deficits: i) An auditory stimulus with rapidly changing elements is not detected / processed in an abnormal way. ii) The speed with which processing of auditory stimuli takes place is abnormal (out of sync). From the literature it is unclear which of these two temporal deficits the ATPDH actually refers to, in fact both can be true at the same time. The early work by Tallal and co-workers suggests the first option (see e.g. Tallal, 1976; Tallal, Miller, & Fitch, 1993; Tallal & Piercy, 1974). However, since the ATPDH has been "adopted" by the magnocellular theory of dyslexia (Stein, 2001; Stein & Walsh, 1997), option two seems more appropriate. This magnocellular theory states that the sensorimotor deficits observed in dyslexic readers may be explained by the anomalies found in the magnocellular neural pathways responsible for fast information transferral. It is thus not exactly clear what the "temporal" in temporal processing refers to. A similar problem plays a role in the rise time perception deficit (e.g. Livingstone, G. D. Rosen, Drislane, & Galaburda, 1991).

The RTPDH states that there are problems with the perception of the slow changing amplitude modulation cues, or rise times of the amplitude envelope of the speech signal. Temporal here thus refers to the opposite of ATPDH in terms of the rate of change involved. The hypothesis has recently been placed in a temporal sampling framework (Goswami, 2011) that provides a neurocognitive basis for the deficit. The main explanatory work in the theory is done by the fact that perceiving changes in amplitude envelopes is essential for segmenting the speech stream into smaller units, for perceiving prosody to mark boundaries of sentences, words and syllables (Ziegler & Goswami, 2005). In one of the first publications presenting this hypothesis (Goswami et al., 2002), it is suggested that the deficit concerns the processing of the acoustic structure of the syllable, which is best, described as rhythm detection. This was tested by asking children to distinguish between stimuli on a continuum from smaller (15 ms) to larger (300 ms) envelope rise times of the modulating wave. The slope of the psychometric categorisation function of the dyslexic readers was smaller than that of typically developing children (compared to chronological age and reading age). The conclusion was that the dyslexic readers were not detecting the envelope onsets that make up the beat of the signal. Performance on the envelope onset detection task explained more variance in reading and spelling performance than the temporal order judgement tasks and rapid frequency discrimination tasks associated with ATPDH. This deficit is also thought to have broader consequences for meter and beat perception in music by dyslexic readers (Goswami, 2006; Huss, Verney, Fosker, Mead, & Goswami, 2010). It is suggested that a deficit in beat perception may also explain why dyslexic readers have problems producing speech, or tapping to a metronome (Corriveau & Goswami, 2009). The causal connection to reading is however still through a deficient representation of a phoneme- like structure due to poor beat perception. This is why the hypothesis belongs in the arena of the component dominant ontology.

The question remains, what exactly is the process that is deficient here? The authors use "rise time perception deficit", "envelope amplitude onset detection deficit", "perceptual insensitivity to amplitude modulation", "beat perception deficit" and "p-centre detection deficit". Recently, the perception of fast spectral changes in formants was directly compared to rise time perception in a /bA/-/wA/ continuum on which stimuli differed either by frequency rate of change or envelope rate of change (Goswami et al., 2010). The frequency onsets of the formants were kept equal in both conditions. It was concluded that dyslexic children were poor at discriminating between sounds based on the rate of change of the envelope, whereas discrimination based on formant transition duration (the rate of change of frequency) yielded normal performance. The authors interpreted the results as a failure to detect envelope cues by dyslexic readers, not rapid frequency changes. What does this imply? Are there too many, or too few rise time onsets in the signal to be perceived. Or, if the deficit is indeed also responsible for anomalous rhythm production, is it a matter of a deficient coupling between an internal clock and an externally perceived rhythm as suggested in the temporal sampling framework (Goswami, 2011)? If those rise time onsets were made more salient, would they lead to better beat perception? Is it a deficit in perceiving the rate with which the amplitude envelope changes in the signal instead of the actual detection of the onset of the envelope? This is what is suggested by the stimuli used in Goswami et al. (2010) and seems a different, more specific auditory processing deficit than the more general deficit the same authors proposed to detect the occurrence of envelope onsets as a beat or rhythm. Most importantly, do dyslexic readers make terrible drummers?

Since confusion about the specifics of the characteristics of the stimuli the deficits pertain to occur in both hypotheses (periodicity or pattern detection vs. rate of change detection), the measures that will be extracted from the speech signal in this study will address both such features. The measures that seem to relate most to a deficient component process appear to be the rate of change of the formant frequency and the rate of change of the amplitude envelope. The periodicity, or pattern measures will be discussed in the next paragraph. To obtain the rate of change of the formant frequency of a stimulus, the Fourier transform of the speech signal is taken and formant tracks are extracted from the spectrum. The slope of the second formant (F2) in the spectrogram is calculated as a measure of rate of frequency change. For RTPDH there are several options to quantify the rate of change of the amplitude envelope. Here, the stimuli used in Goswami (2009) are considered to decide on an appropriate measure, because in that study several of the options mentioned above (rise time duration, envelope onset, tempo, etc.) are contrasted against one another. Differences between dyslexic readers and typical readers (chronological age controls) in that study were significant when discriminating between two types of stimuli: stimuli with single ramp envelope onsets (with random steady states and rise times varying from 15 to 300ms) and composite stimuli consisting of a standard rise time (15 ms) alternated with a longer rise time (up to 192 ms). The study showed that performance on discrimination tasks with these stimuli was correlated with rhyme detection and reading and had a unique contribution to explained variance in these variables in a regression model. A sensible measure then seems to be the

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Figure 1. Figures represent spectrograms of the stimuli used in the experiment. There are four manipulations of the 10-step continuum. Formant Sweeps (Δ F2) are calculated as the slope of the second formant transition (the second white line).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

time it takes for the amplitude envelope to rise to its maximal value, which also marks the onset of the rhyme (b-Ak). To obtain the measure, first the Hilbert transform of the signal is taken (cf. Feldman, 2008; Smith, Delgutte, & Oxenham, 2002), which yields the immediate envelope. The slope of the line one can draw from the amplitude envelope at the start of the signal to its maximum value is considered an estimate of the most important slow rise time that needs to be detected in order to distinguish between speech stimuli.

Acoustic manipulations of the speech signal, based on the ATPDH refer to amplification or slowing down (or both) of the fast spectral changes present in the speech signal. These manipulations are expected to also affect the amplitude envelope, which is important for RTPDH. Amplification may lead to steeper rise time slopes whereas slowing down the signal is expected to lead to (relatively) slower rise times. Been and Zwarts (2003) presented simulations of the effect of amplification of the fast formant transitions using their SWEEP model. The SWEEP model is a dynamical model built around the assumption that speech perception involves detection of frequency sweeps. They predicted that the amplification manipulation would indeed lead to a better performance on behalf of the dyslexic readers. Following this line of reasoning, we may expect a measure that indexes the rate of change of a formant transition in a speech signal to be a measure of which ATPDH would agree dyslexic readers cannot maximally exploit to identify and discriminate between speech sounds.

In Figure 1 the spectrograms of the stimuli used in the present study are plotted. The rate of change of the formant transition calculated as the slope of F2 in the spectrum is given for each of the 40 stimuli. Figure 2 shows the smoothed amplitude envelopes of all the stimuli and the rise time is calculated as the slope from the start of the stimulus to the maximum amplitude. As shown in the figure, these measures differ between the stimuli and are thus candidate features that may actually be used by participants.

3. Periodicity Measures: Harmony of Frequency and Amplitude

The periodicity measures used are Rise- and Fall-Time Entropy (RFTe) and Inharmonicity (also known as Harmonics-to-Noise-Ratio, HNR). In theory these measures should be connected to RTPDH and ATPDH respectively. Quite remarkably to my knowledge, they have never been used in studies in the context of speech perception and developmental dyslexia.

RFTe represents the entropy (disorder) in the distribution of rise and fall times estimated present in the envelope. It is calculated by taking the first derivative of the immediate amplitude envelope (obtained by a Hilbert transform of the signal), which represents the rate of change of the amplitude. When the differenced amplitude envelope changes sign, that is crosses the x-axis, there is a peak in the amplitude (rate of change is zero) after which the amplitude rises or falls. Quantifying the time between peaks in the envelope by subtracting the timestamp of subsequent zero-crossings in the derivative thus

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

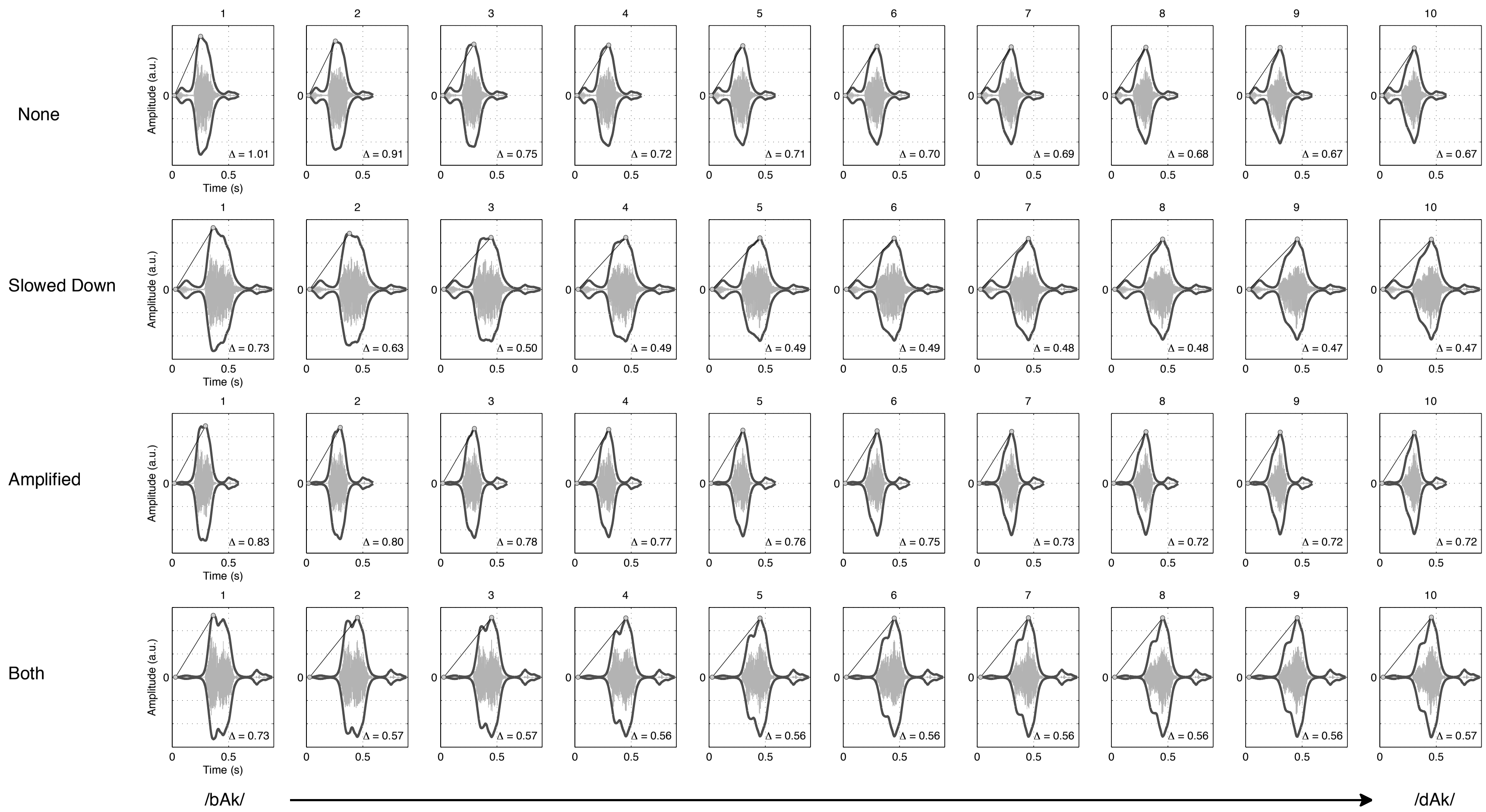

Figure 2. Figures represent the smoothed envelope (exaggerated for clarity of presentation) of the amplitude waveform (smaller ghost image) of the stimuli used in the experiment. There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. Envelope rise times (Δ) are calculated as the slope of the line connecting the start of the stimulus onset amplitude to the maximum amplitude.

yields a distribution of durations; the time it takes for the amplitude to rise or fall. The entropy of this distribution of discrete durations of size *n* can be calculated as the chance of observing a particular rise or fall time $pRFT_i$ (equation 1) and inserting it into the regular formula for Shannon entropy (equation 2).

$$pRFT_i = \frac{RTF_i}{\sum_{i=1}^{n} RFT_i} \tag{1}$$

$$RFTe = -\sum_{i=1}^{n} pRFT_i * log_2(pRFT_i) \tag{2}$$

RTFe may be considered an estimate of the harmony of the perceptual rhythm invoked by amplitude changes. High entropy means that there is disorder or noisiness in the amplitude envelope of the signal. Another way to interpret entropy is in terms of information: The value of the entropy denotes how many bits (because log2 is used) of information would be needed to predict the rate of change of the envelope. More bits needed means less regularity and more disorder in the curve. The RTFe values for each stimulus are shown in Figure 2. The figure reveals RFTe takes on different values for different steps on the continuum, but also across different acoustic manipulations.

Table 1

Inharmonicity of the 40 Stimuli Used in the Experiment. The Numbers Represent Percentage of Energy in the Signal that is Outside of the Harmonic Sequence.

		Acoustic Manipulation			
		None	*Slowed Down*	*Amplified*	*Both*
stimulus	*/bAk/*	39.87	42.52	43.81	48.79
	2	38.99	42.50	42.80	48.57
	3	38.40	41.09	42.38	46.26
	4	37.46	41.01	42.32	47.30
	5	37.09	40.58	42.24	46.45
	6	36.95	40.37	42.14	46.05
	7	36.81	40.13	42.00	45.13
	8	36.85	39.67	40.70	414
	9	36.71	39.57	40.59	43.84
	/dAk/	36.20	38.89	40.91	43.35

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

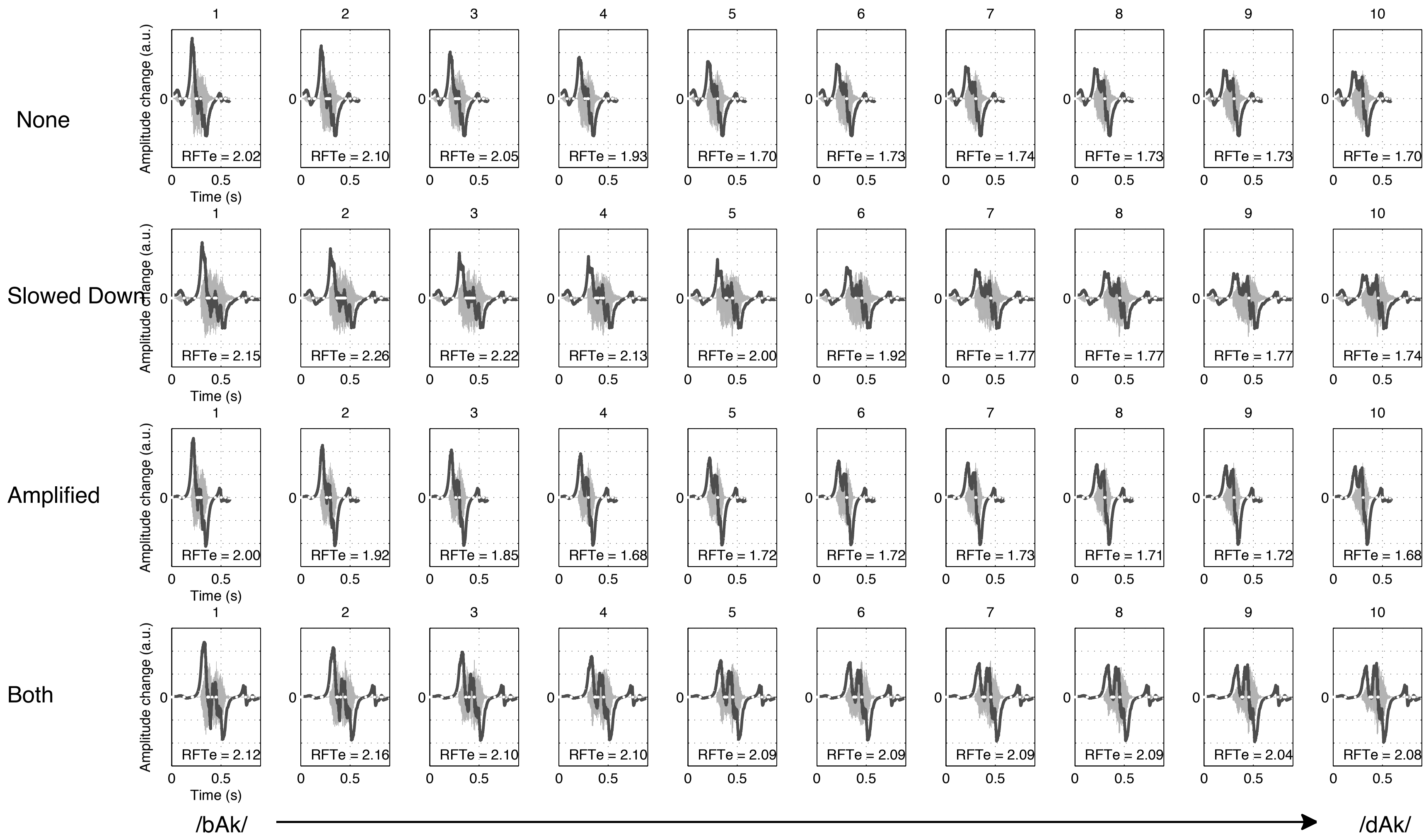

Figure 3. Figures represent the derivative of the smoothed envelope (exaggerated for clarity of presentation) of the amplitude waveform (smaller ghost image). There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. The derivative represents the rate of change of the envelope. The dots are the zero-crossings of the derivative and the time between consecutive dots represents a rise or fall time. See text for details on calculating RFTe values.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Inharmonicity, or HNR measures how much energy in the spectrum is outside of the ideal harmonic sequence. To calculate this measure we assume the signal may be decomposed into a large number of partials, or sine waves that oscillate at a particular frequency. We also assume there is a fundamental frequency F0. The more harmonious the signal, the more it consists of partials that are multiples of F0. The Formants discussed earlier, can be considered such multiples. In an ideal situation, the second formant frequency F2 should be 2^n*F0, with n=2. For the calculations presented here, the exact correspondence of the value of n to the order of the formant is not important as long as it is a multiple. Inharmonicity then represents how many of the partials in the signal are not multiples of F0, how much the signal deviates from an ideal harmonic sequence. This measure captures information about the impact of the changes in formant frequency with respect to the other formant frequencies present in the signal and might be a more accurate index of spectral changes than the absolute change in one formant such as the F2 slope. Table 1 lists the inharmonicity values of the stimuli as the percentage energy in non-harmonic partials. Again, there are clear differences between stimuli on the continuum and between the acoustic manipulations.

The stimuli used were synthesised (but based on actual recordings of utterances, see van Beinum, Schwippert, Been, van Leeuwen, & Kuijpers, 2005) to create a continuum in which the F2 onset frequency is the only major spectral change. F2 is constant at 1100 Hz in /bAk/ but the onset increases in ten steps to 1800 Hz in /dAk/. In the table it can be seen that /bAk/ is more inharmonious than /dAk/, which might seem counterintuitive since in /bAk/ there is no change in F2 onset. However the fundamental frequency F0 of most of the stimuli is about 220 Hz, which yields about 1800 Hz with n=3. The closest harmonic partial to 1100 Hz is 880 with n=2.

4. Interaction Dominant Dynamics Measures: Quantifying Complex Temporal Patterns

A spectrogram representation of a speech sound (see Figure 1) reveals the complexity of the speech signal by displaying how much the energy at different frequency bands changes over time. Since the stimuli used here are partially synthetic, the spectrograms presented in Figure 1 are less noisy than recordings of actual speech produced by a human would be. When trying to understand how humans perceive such a signal as a meaningful word or sentence, it is tempting to focus on mechanisms that analyse frequencies or amplitudes and loose sight of the fact that the spectrogram is a representation of a complex gesture, a motor action. In fact, there are at least 70 muscles involved in producing even a simple syllable like /pa/ ranging from muscles that control respiration to the ones that control the tongue (Galantucci, Fowler, & Turvey, 2006; Turvey, 2007). Producing speech sounds is very much a matter of sophisticated aerodynamic control by changing the shape of cavities air is forced to flow through (Porter & Hogue, 1998). Models of aspiration for example have been successfully validated against real turbulent airflow induced sound, generated in acoustic duct experiments (Little et al., 2007). Fasten your seat-belts; hold on to your coffee! Human

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

speakers are able to create turbulence at will…

Several authors have suggested aggregate, or collective levels of control that enable us to understand a control task with mind-boggling numbers of degrees of freedom. The uncontrolled manifold (Scholz & Schöner, 1999) and synergies (Turvey, 2007) are examples of such higher order mechanisms of control. They represent theoretical constructs based on the interaction dominant ontology in which the interactions between all the components and the way larger component ensembles are coupled to do the explanatory work for the theory. The most sophisticated theoretical frameworks treat action and perception as a coupling of levels in a single complex system whose behaviour needs to be explained as an inseparable whole (e.g. Chemero, 2009; Chemero & Turvey, 2007; Gibson, 1979; Michaels & Carello,

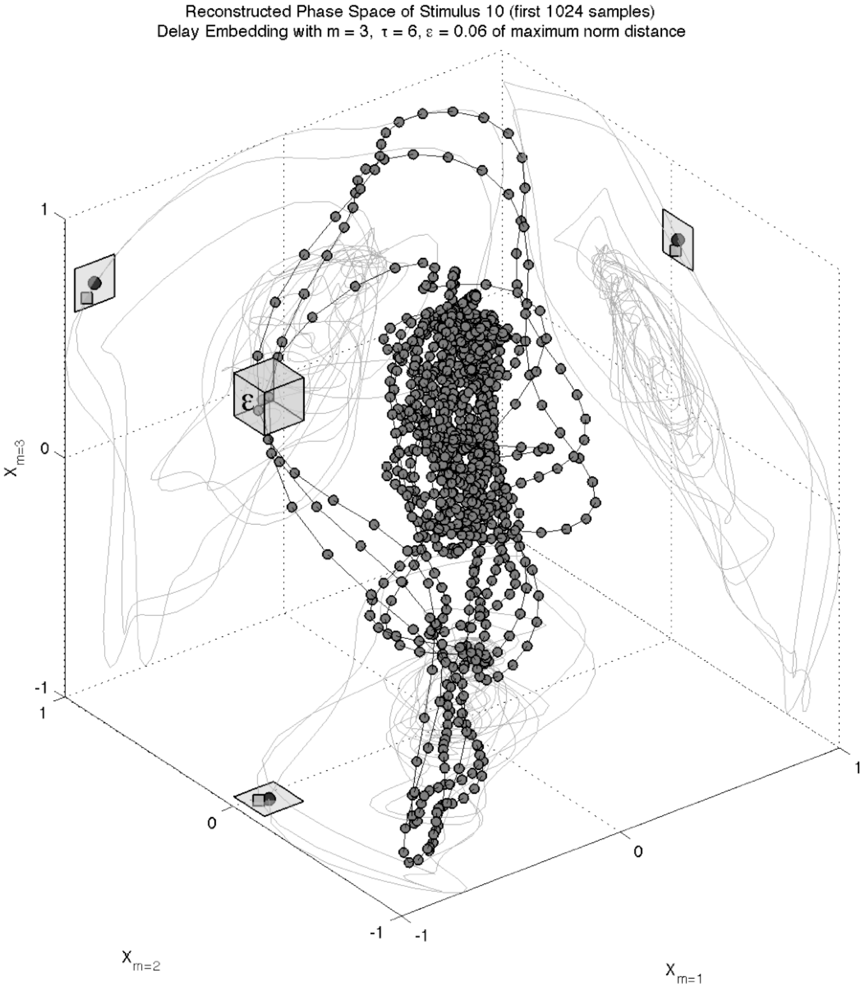

Figure 4. A reconstruction of the 3D phase space of stimulus 10 (first 1024 samples) by the method of delay embedding. The planes show 2D projections of the time course of the surrogate dimensions created with an embedding delay $\tau = 6$. Points that fall within a distance ε (represented by the grey box for presentation purposes) will be plotted as recurrent points in the recurrence plot.

1981; Schöner & Kelso, 1988). Most theories of speech perception, however, seem to ignore the action perspective and focus on sound as a physical signal, of waves propagating through

air that need to be analysed for their frequency and amplitude characteristics (Porter & Hogue, 1998). Perhaps we should listen to the words of Horace Lamb, the author of the 1910 book: *The Dynamic Theory of Sound*, which is still in print today as an exact copy of the 1925 2nd edition (Lamb, 2004). He was more famous for his work in hydrodynamics and is reported to have said: "*I am an old man now, and when I die and go to heaven there are two matters on which I hope for enlightenment. One is quantum electrodynamics, and the other is the turbulent motion of fluids. And about the former I am rather optimistic.*" (Tabor, 1989, p. 187). The founding father of quantum electrodynamics, Richard Feynman called turbulence "*the most important unsolved problem of classical physics*" (Moin & Kim, 1997). What becomes quickly apparent when the dynamic theory of sound is examined is that a speech sound is not the same as the vibration of a violin string propagating harmonic waves through the air. The sound wave produced by a string is to the sound wave produced by a human speaker as a gentle summer breeze is to a hurricane.

Turbulence can be observed in any propagating medium and may be (partially) described as spatiotemporal chaos, or deterministic randomness in time and space simultaneously. As a consequence it is very difficult to accurately measure, model, forecast, or control turbulence in a medium. Methods have been developed however to analyse or quantify turbulent dynamics, most notably by exploiting the theorem proved by the Dutch mathematician Floris Takens in a lecture note entitled "*Detecting strange attractors in turbulence*" (Takens, 1981). Takens' theorem states that the m-dimensional attractor of a dynamical system may be reconstructed from a measured time series of a single observable dimension of that system. The idea is that since the behaviour of the system is governed by interactions on many different spatial and temporal scales (interaction dominant dynamics), information about the dynamics of the whole system is present in the dynamics of its parts (the measured dimension). By using m delayed copies of the observed time series as surrogate dimensions that span the state space in which the dynamics of the entire system unfold we can reconstruct its attractor dynamics. Takens' theorem ensures that the reconstructed attractor is topologically equivalent to the actual attractor when all of the m dimensions of the system are indeed observed (see Marwan, Carmen Romano, Thiel, & Kurths, 2007 for an explanation)

After phase-space reconstruction the next step in the analysis of a highly interaction dominant system is to quantify the dynamics of the reconstructed attractor. A method commonly employed for this purpose is Recurrence Quantification Analysis, RQA (Marwan et al., 2007; Webber Jr., Marwan, Facchini, & Giuliani, 2009; Zbilut, Giuliani, & Webber Jr., 1998; Webber Jr. & Zbilut, 2005). RQA is a nonlinear time series analysis technique that can quantify complex temporal patterns by means of analysing trajectories through state space and noting when trajectory coordinates appear in each other's vicinity. In Figure 4 the attractor of the first 1024 samples of the transition part of the amplitude time series of stimulus 10 (/dAk/) is reconstructed in three dimensions. The time series for surrogate dimension m is shifted by τ samples for each extra surrogate dimension m. The values for τ and m are chosen so that the reconstructed attractor will represent maximal information

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

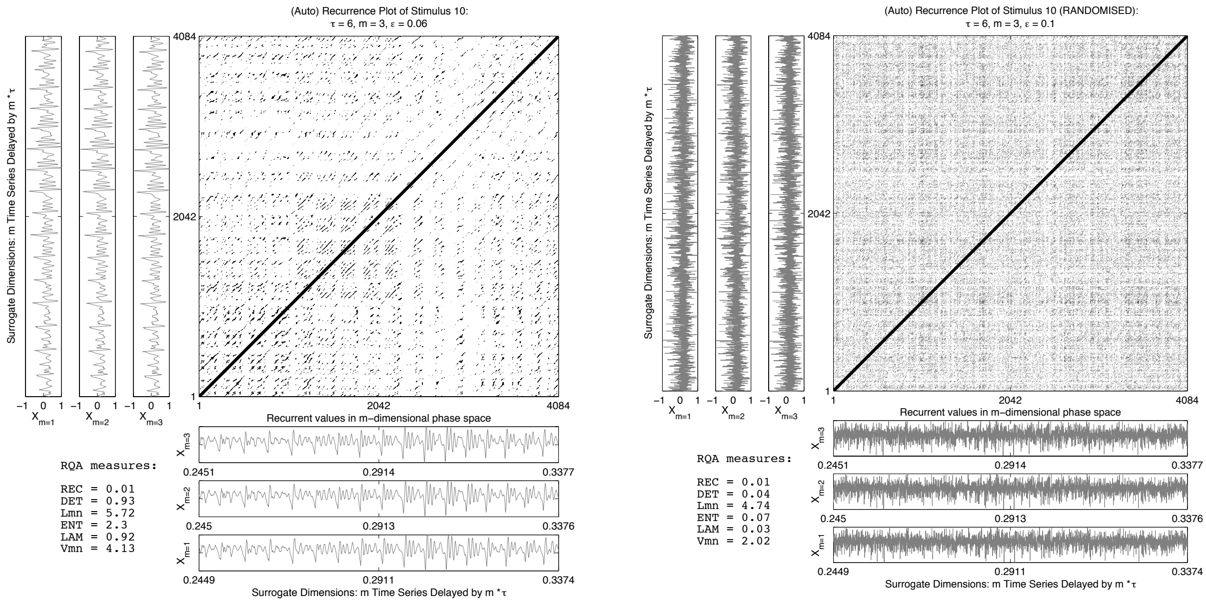

Figure 5. A recurrence plot of the transition part of stimulus 10 (left) and a randomly shuffled version (right). Also plotted next to the recurrence plots are the surrogate dimensions m that span the phase space in which recurrent points are evaluated. They are offset by just (m-1)*6 samples. The recurrence quantification measures show the effect of randomising the temporal order of the series: The recurrence rate is the same, but the recurrent points do not form line structures in the randomised version.

present in the measured series (mutual information is used to choose τ and a false nearest neighbour analysis to choose m, see Riley & Van Orden, 2005 for details). The coordinates in reconstructed state space in Figure 4 are not randomly jumping from one region to another, but trace periodic orbits through specific locations in the state space. When two coordinates fall within a radius ε the two coordinates are said to be recurrent. Sequences of multiple coordinates that are recurrent signify a trajectory in phase space that is being revisited by the system. It is a trajectory or a location in the state space the system is attracted to and these recurrent coordinates and the structures they form are the objects of analysis in RQA.

In Figure 4 trajectories are clearly visible as orbits around the denser centre of the state space. It is also apparent that the choice for a radius size will greatly influence which coordinates will be recurrent (e.g. Schinkel, Dimigen, & Marwan, 2008). In general the radius, or threshold used in RQA is set to a number that yields 1-5% recurring coordinates (out of all theoretically possible recurring points given the size of the state space). The recurrent coordinates are recorded in a recurrence matrix visualised by a recurrence plot of which an example is shown in Figure 5. Since we are looking at recurrent trajectories of one system the time series of m-dimensional coordinates is evaluated against itself (auto-recurrence). For each coordinate pair a distance can be established and if that distance is smaller than the radius a black dot is plotted. The dot represents the fact that at some point in time the coordinate under consideration will be revisited by the system, approximately that is. This yields a recurrence plot that can contain horizontal and vertical line structures as well as individual recurrent points. Diagonal line structures represent a sequence of different coordinates (a trajectory through state space) that is revisited by the system, the proportion of

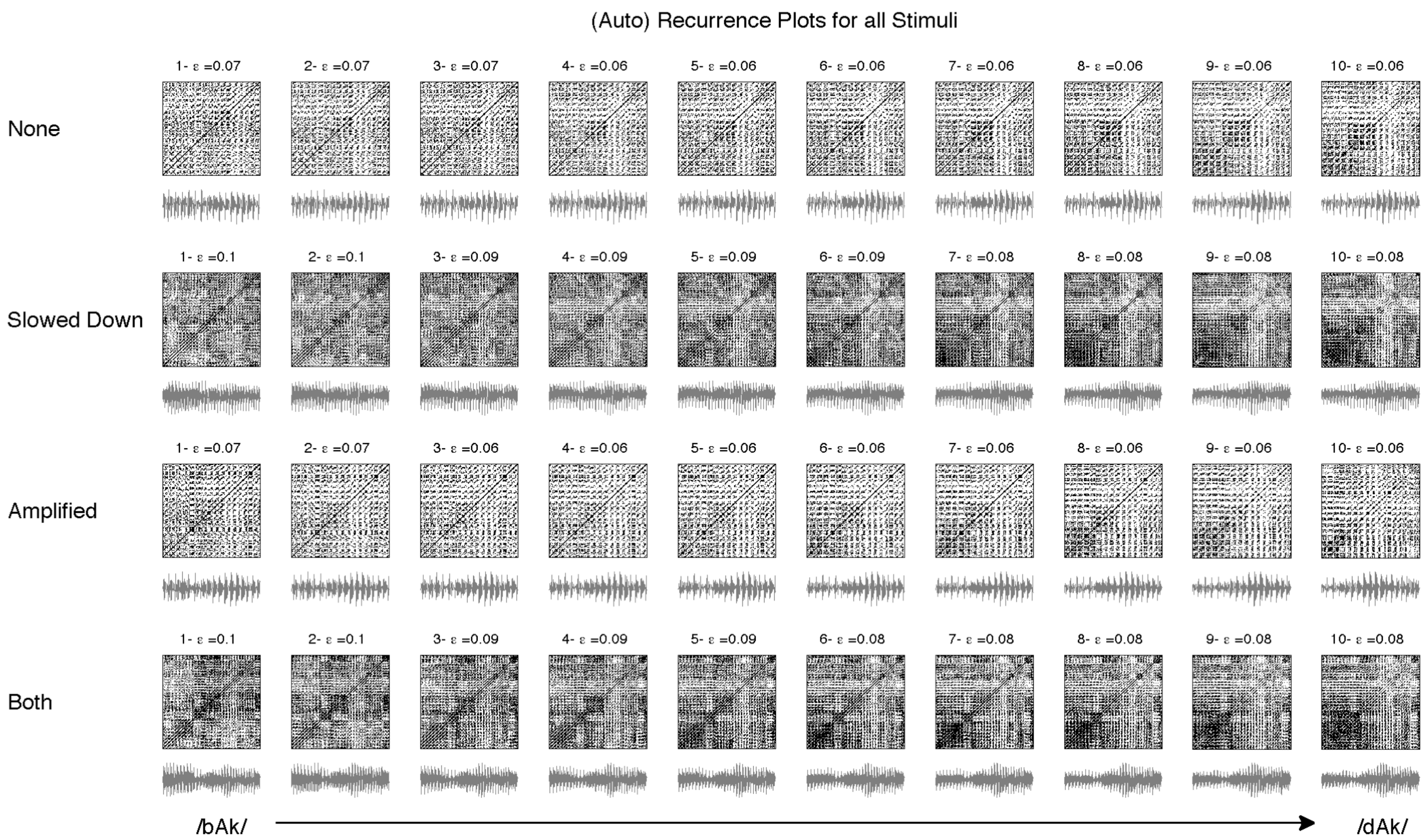

Figure 6. Figures represent the recurrence plots of the amplitude waveform of the transition part of the stimulus (grey image below the plot). There are four manipulations of the 10-step continuum from /bAk/ to /dAk/. The plots were generated using embedding dimension (m) of 3 and an embedding delay (τ) of 6. The recurrence rate for each plot was kept constant by varying the radius (ε). This way the recurrence measures extracted from the plot are comparable across stimuli. See text for details.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

recurrent points that form a diagonal line is quantified as determinism. A vertical line structure signifies that system dynamics are attracted to a specific location in state space where it remains for a longer period of time. The proportion recurrent points that form a vertical line is called laminarity and the mean vertical line length is called trapping time. One could say it quantifies whether the dynamics get 'trapped' in some region of the state space for a while. The plot is symmetrical about its diagonal, which represents the line of identity, or line of temporal incidence. By definition this is the longest line structure in the plot and is excluded from calculations.

The different line structures are clearly visible in the left pane of Figure 5 that is the recurrence plot of the entire reconstructed phase space of the transition part of stimulus 10, the first 1024 samples of which are shown as a reconstruction in Figure The right pane of Figure 5 is a randomised version of stimulus 10, the temporal order of the samples was randomised, destroying all the correlations that are in the data but retaining the same distributional properties (mean, variance, etc.). From the recurrence measures it can be seen that the recurrence rate in both panes (the number of recurrent points) is exactly the same. However, the measurers that are calculated from the line structures that quantify the higher order recurrent patterns are very different. In the randomised plot all the determinism and laminarity disappeared, the temporal structure was destroyed even though the central tendency measures are exactly the same. This is a very basic test of whether the line structures are just accidental temporal alignments. A more sophisticated test would be to create spectral surrogates of the speech stimuli, or to do a bootstrap resampling on all the recurrence measures in order to create a confidence interval (cf. Schinkel, Marwan, Dimigen, & Kürths, 2009). Figure 6 shows the recurrence plots for all the stimuli used in the present study. The threshold was varied in order to keep the recurrence rate exactly the same for all stimuli under consideration. Since we are looking at recurrences in reconstructed phase space, the assumption is that the figures represent the dynamical behaviour of the complex system that produced the speech signal.

RQA is not as exotic as it might appear when first introduced to the technique. In the social and life sciences alone it has been applied in an increasing number of studies across the different sub-disciplines such as motor development in infants (Assmann, Romano, Thiel, & Niemitz, 2007), parent-child interaction (De Graag, Cox, Hasselman, Jansen, & De Weerth, 2012; Lichtwarck-Aschoff, Hasselman, Cox, & Granic, 2012), syntactic coordination between child and caregiver (Dale & Spivey, 2006), dynamics of motor control (Diniz et al., 2011; Wijnants, Bosman, Cox, Hasselman, & Van Orden, 2011; Wijnants, Bosman, Hasselman, Cox, & Van Orden, 2009), cognitive constraints on postural stability (Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Santana, & Fowler, 2003), eye-movements during conversation (D. C. Richardson, Dale, & Kirkham, 2007), insight in problem solving (Stephen, Dixon, & Isenhower, 2009), and as a novel analysis tool in cognitive neuroscience (Bianciardi et al., 2007; Schinkel, Marwan, & Kürths, 2007, 2009). The quantification measures as used in these studies often take their role as predictors or covariates next to more familiar variables in some statitsical analysis. Such will be the case in this study where

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

determinism (DET) and laminarity (LAM) of the recurrence plots of the stimuli (Figure 6) will serve as features to classify the stimuli. DET and LAM are obtained by reconstructing the phase space of the transition part of the stimuli that were resampled to have the same number of samples. Values used for reconstruction were m = 3 and τ = 6 and the recurrence rate was kept constant by varying the radius ε (radius values are shown in Figure 6; DET and LAM values are shown in Table 2). As explained above, DET quantifies recurring trajectories through phase space and a high DET signifies a system that behaves very periodic and predictable. LAM quantifies recurrences of the system displaying the same type of behaviour, visiting the same region in phase space and staying there for a while. Some portion of the recurrent points quantified by DET will be representing laminar behaviour, so using a combination of these two measures in a classification analysis yields a description of the stimulus in terms of whether the dynamics are characterised by changing temporal patterns or patterns that stay relatively constant for some time.

5. Which measure do participants use to identify /bAk/ and /dAk/?

A recent successful application of RQA and other complexity measures to speech sound classification was done in the context of voice disorder detection (Little et al., 2007). Natural recordings from a database of more or less clear examples of voice disorders were analysed on the classification ability of several measures thought to be theoretically important to detect the voice disorders (jitter, shimmer, amplitude irregularity, and HNR). These classical measures, together with the complexity measures Recurrence Period Density Entropy (RDPE, a measure derived from the recurrence times in the plot) and a normalised scaling exponent (H_{norm}, derived from Detrended Fluctuation Analysis; DFA) were evaluated for their classification performance in a quadratic discriminant analysis (QDA). The complexity measures were superior in distinguishing between normal and voice disorder recordings (overall classification 91.8% correct for RDPE/H_{norm} with other measure pairs ranging from 76.4% to 81.4%; see Table 1 in Little et al., 2007).

In this study I will use a similar approach to categorise the speech signals as Little et al. (2007) did, but the targets for the quadratic discriminant analysis (QDA) will not be disordered speech vs. healthy speech, but the observed labelling of the stimuli by average and dyslexic readers as either /bAk/ or /dAk/. The labelling patterns will be experimentally assessed by administering a labelling task of four versions of a 10 step /bAk/ to /dAk/ continuum (None, Slowed Down, Amplified and Both). A first research question is wether there are differences in labelling between experimental groups and stimulus types. This could potentially yield eight different labelling patterns. If there is a difference between experimental groups, QDA will be performed for each group separately. The features used by QDA to classify the stimuli will be the measures discussed above. These measures are extracted from one and the same set of stimuli, but represent different theoretical perspectives on (impaired) speech perception. The simple main hypothesis is that the combination of measures that yield the best classification performance is the most likely source of

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Table 2
Determinism and Laminarity of the 40 Stimuli Used in the Experiment. The Numbers Represent Proportion of Recurrent Points That Lie on Diagonal Lines (DET) or on Vertical Lines (LAM).

Stimulus	*Acoustic Manipulation*							
	None		*Slowed Down*		*Amplified*		*Both*	
	DET	LAM	DET	LAM	DET	LAM	DET	LAM
/bAk/	0.95	0.91	0.90	0.83	0.95	0.90	0.86	0.78
2	0.95	0.91	0.89	0.82	0.95	0.91	0.86	0.78
3	0.95	0.92	0.89	0.83	0.95	0.91	0.85	0.78
4	0.94	0.91	0.88	0.82	0.94	0.91	0.84	0.77
5	0.94	0.92	0.87	0.81	0.94	0.91	0.83	0.77
6	0.94	0.91	0.86	0.81	0.94	0.91	0.82	0.77
7	0.94	0.92	0.85	0.81	0.94	0.91	0.81	0.78
8	0.94	0.92	0.84	0.81	0.93	0.91	0.80	0.78
9	0.93	0.92	0.83	0.81	0.93	0.91	0.79	0.77
/dAk/	0.93	0.92	0.82	0.81	0.94	0.91	0.77	0.76

information used by the participants in this study to label the stimuli.

6. Method

Participants

Participants in this study were recruited based on the response of parents of dyslexic children to information distributed at local schools and printed in local newspapers. The average reading children were recruited from the same classroom or school as the participating dyslexic children. Participation of all the children in the experiment was based on consent provided by the parents. This resulted in a sample of 88 children (age range 101.2 to 159.3 months) from 9 different schools in the southeast of the Netherlands. Half of the subjects (44) were dyslexic readers as indicated by two reading tests: A timed-reading task for regular words ("Drie-Minuten-Toets"; Verhoeven, 1995) and a timed pseudo-word reading task ("KLEPEL"; van den Bos et al., 1994). When the child's scores on both tests were within the 25th percentile (norm score by age), the child was considered to have moderate to severe reading problems. See Table 3 for details.

Stimuli and Acoustic Manipulations

The stimuli were based upon natural speech recordings for the words /bAk/ [container] and /dAk/ [roof] and transformed to create a 10-step /bAk/ to /dAk/ continuum (van Beinum, Schwippert, Been, van Leeuwen, & Kuijpers, 2005) using the Praat program (Boersma & Weenink, 2001). The stimuli differed only with respect to the second formant

transition of which the onset frequency was gradually increased from /bAk/ to /dAk/ (see Table 5 for exact values).

All the stimuli on this F2 continuum were manipulated in three manners using the Praat program (Boersma & Weenink, 2001). First, the speech signal was *Slowed Down* to 150% of its original length. This was achieved by a Pitch Synchronous Overlap and Add (PSOLA) algorithm (see e.g. Segers & Verhoeven, 2005). Second, the signal was *Amplified* by 20 dB, for the fast changing spectral elements. The algorithm used to do this in Praat was similar to the one used by Nagarajan (1998), who confirmed this in a personal communication with Segers and Verhoeven (2002). Third, *Both* manipulations were applied as is done in the FastForWord program (Merzenich, et al., 1996; Tallal, et al., 1996): the speech signal was slowed to 150% of its original length and all the fast transitional elements were then amplified by 20 dB. There was of course also a continuum which had *None* of the manipulations applied to it. This yielded 40 different stimuli in total.

Table 3

Results For the Two Groups of Children Participating in the Experiment. The DMT Scores Represent Words Read Correctly in One Minute. Level Of Difficulty Increases From DMT1 to DMT3. KLEPEL Represent Correctly Read Pseudowords In Two Minutes.

	Average Readers		*Dyslexic Readers*	
	Mean	SD	Mean	SD
Age (months)	128.1	12.3	133.2	12
DMT1	99.9	15.5	73.4	15.3
DMT2	95	18.0	61.1	15.3
DMT3	83	16.2	49.0	16.1
KLEPEL	71	17.4	33.5	12.5
Gender	21 Boys	23 Girls	23 Boys	21 Girls
N	44		44	

7. Procedure

Speech perception experiments

The speech identification task (labelling task) was presented on a laptop computer in a quiet room at the children's school. There were two tasks conducted in two sessions; an identification task (reported in this study) and a discrimination task (reported in Hasselman, 2014c). In the identification task, the participants were presented a smiley face on the screen, which then uttered a word. After utterance of the word, two frames with a picture in each appeared on the left and right of the screen. The pictures in the frames were either a roof or a container. The pictures presented in the frames were randomly interchanged at each

presentation. Prior to the experimental trials, 10 practice trials were presented using different words and pictures that were al clear examplars, allowing feedback on their response. During the experimental condition, the three types of manipulated /bAk/ and /dAk/ stimuli and the unmanipulated stimuli were presented in a random order. Each stimulus was presented twice resulting in 80 stimulus presentations (2 x 4 manipulations x 10 stimuli).

Extracting the stimulus Characteristics

The 40 stimuli were 16 bit digital audio files in .WAV format, with a sample rate of 41 KHz.[1] These were always used as the basis for extracting the following measures: The slope of the second formant transition (F2 slope, Figure 1), the time it took for the envelope to reach its maximal value (mxENV Slope, Figure 2), and the entropy of rise and fall times (RFTe, Figure 3). Extraction of these measures is described in detail in Appendix 1. which includes all the MATLAB code (version R2010b; The Mathworks Inc., 2010) necessary to perform the calculations. Settings were used in MATLAB that mimic the default behaviour of the Praat program (Boersma & Weenink, 2001), so the output of this script should be similar to output generated by Praat. For the Inharmonicity measure (HNR; Table 1) and the measures obtained from recurrence quantification analysis (Figure 6) only the transition part of the stimulus was considered. Following Little et al. (2007), to assure that the RQA is performed on time series of equal length, all files were resampled to 4096 samples (waveforms shown under the RP plots in Figure 6). The MATLAB code to create the figures and perform the QDA analysis is available in Appendix 3 and

Statistical Analysis

For each participant there were 80 responses of either /bAk/ or /dAk/. These data were entered in a logistic multilevel model (using MLwiN version 2.2; Rabash et al., 2009) with the 80 measurement occasions representing responses to a random permutation of the ordered F2 continuum at level1. The responses at the level of the measurement occasions were considered binomially distributed as 0 and 1 and a logit link function was used. The repeated measurements can be thought of as clustered within the participant, who represent a second level of random variation in the model (level2). The modelling strategy was as follows: First it was examined whether the multilevel model gave a better fit than a single level model with just measurement occasion defined as a level. Then, the empty multilevel model for change was fitted (M0), which in the present case means that a zero inflated fixed effect predictor was added representing the stimulus rank order on the continuum (0-9). In a subsequent model (M2) it was examined whether stimulus rank could explain random variation in the slopes of the curve at the level of the participants (level2). If so, this means the variation in labelling of the continuum between participants can be understood as random variation with

[1] All data and MATLAB scripts used in this study are available online at the open science framework project Beyond the Boundary - Chapter 4: http://osf.io/2r5eh

respect to the average labelling curve of the entire sample. In the next step (M3) level1 and level2 covariates were added: A dummy variable that represents the four stimulus types (level1), and a dummy variable that represents whether subjects are dyslexic or average readers (level2). In the final modelling step (M4) various interactions were tested including cross-level interactions between participant type and stimulus type. The models were fitted using a Monte Carlo Markov Chain (Brown, 2009) simulation with 150,000 iterations. This

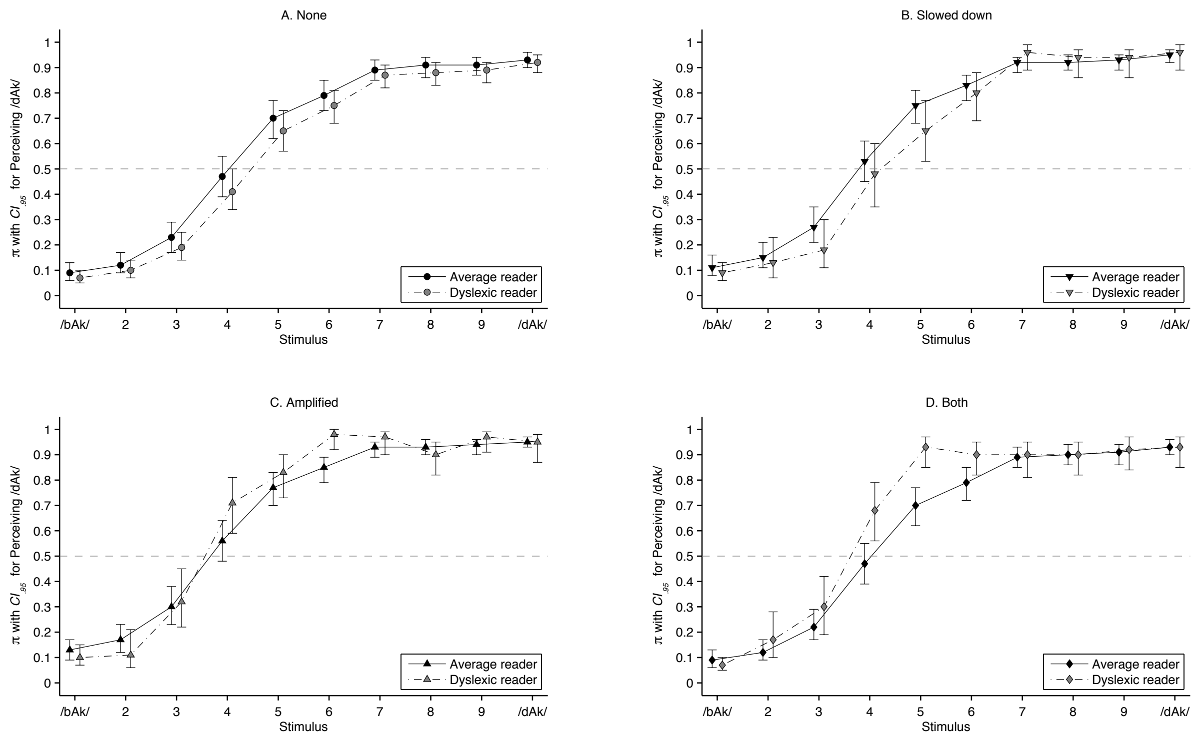

Figure 7. Predicted probability of perceiving /dAk/ with 95% CI for each stimulus on the artificial continuum (predictions based on 150,000 MCMC replications of model M4 in table 5). The lines summarise average and dyslexic readers and panels represent each type of manipulation: A. *None*; B. *Slowed Down*; C. *Amplified*; D. *Both*. The points are offset around the stimulus number on the x-axis to increase readability. There are two clear instances of non-overlapping confidence intervals (Panel C, stimulus 6; Panel D, stimulus 5). Values for the entire sample (Model M3) are given in Table

number was chosen after inspecting the Raferty-Lewis diagnostic for each parameter estimate at each modelling step and was found to yield a very safe margin for all predicted parameters.

The predictions of the logistic multilevel model for each stimulus were used as targets for the quadratic discriminant analysis (QDA). If the lower 95% confidence bound predicted by the logistic multilevel model exceeded the chance level of 0.5 it was noted for that stimulus that /dAk/ was perceived. Otherwise the target for the discriminant analysis was /bAk/ for that stimulus. This resulted in a string of 40 zeroes and ones. The objective of the discriminant analysis was to replicate the classification in zeroes and ones based on pairs of the measures discussed above. The following pairs were tested mxENV Slope / F2 Slope; HNR / F2 Slope; RFTe / mxENV Slope; RFTe / HNR; LAM / DET. The pairs were all

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

converted to the unit scale before analysis. The algorithm used to perform QDA was the same as described in Little et al. (2006). This procedure allows for calculation of 95% Confidence Intervals around the percentage correctly classified stimuli by bootstrap resampling. All QDA analyses were based on 15,000 bootstrap replications. Details about the procedure for applying QDA and the creation of Figure 8 in this study are available as MATLAB code in appendix

8. Results

Multilevel Logistic Model

The results of multilevel modelling taking the individual trials of the identification experiment as the dependent variable at level1 and subjects at level2 are shown in Table A graphical representation of the predictions by the final model is shown in Figure 7. In the final model, there was no significant main effect of experimental group (dyslexic reader vs. average reader), but there was a significant cross-level interaction between experimental group and acoustic manipulation. This interaction is revealed in Figure 7 where in Panel C (Amplified) and D (Both) there two clear examples of non-overlapping CI between the labelling curves of average and dyslexic readers for stimulus 6 in Panel C and stimulus 5 in Panel D.

In both cases the dyslexic readers have a higher odds for perceiving /dAk/. Another difference between the groups may be observed when evaluating at which stimuli the lower confidence bound of the odds for perceiving /dAk/ exceeds the chance level of 0.5. Again the difference between the groups is observed with stimuli of category Amplified and Both (Panel C and D in Figure 7). The dyslexic readers' odds for perceiving /dAk/ is with 95% certainty higher than chance at stimulus 4 for these manipulations whereas for normal and Slowed Down manipulations it is at stimulus 5. For average readers this boundary is always at stimulus 5 irrespective of the acoustic manipulation. In Table 4 the significant parameter estimates of the final model (M4) corroborate this: At each unit step increase in F2 frequency (stimulus number) there is an increase in the odds of perceiving /bAk/. Amplified stimuli also increase the odds of perceiving /dAk/ and for the group of dyslexic readers Amplified and Both stimulus types add even more to those odds. The random intercept and slope variance indicate that labelling curves vary across participants. Adding predictors and cross-level interactions did however not noticeably decrease, or explain this variance (changes are in 3rd decimal). The DIC statistic did decrease with each consecutive model indicating a better model fit.

Quadratic Discriminant Analysis

Because the outcomes of the multilevel logistic model yield different boundaries at which dyslexic and average readers switch from /bAk/ to /dAk/ for stimuli of type Amplified and Both, the QDA was performed for each group separately using these

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Table 4

Model Evaluation With Identification Label (idL) As Dependent Variable. The Bayesian Deviance Information Criterion Was Used For All Consecutive Models Estimated with MCMC (150,000 iterations). D = Posterior Mean Deviance, D(ϕ) = Deviance of Posterior Means, pD(D-D (ϕ)) = Effective Number of Parameters, DIC = Deviance Information Criterion. Significant Effects of M4 Printed Italic. See Text For An Explanation of the Modelling Steps.

	Msingle		*M0*		*M1*		*M2*		*M3*		*M4*	
idL_{ij} =	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.	β	S.E.
Fixed Part												
Intercept	0.51	0.03	0.53	0.04	-2.04	0.10	-2.39	0.15	-2.77	0.19	-2.59	0.20
stimulus					0.66	0.02	0.77	0.04	0.8	0.05	0.78	0.04
Slowed Down (D1)									0.29	0.10	0.16	0.14
Amplified (D1)									0.63	0.10	0.37	0.14
Both (D1)									0.36	0.10	0.01	0.14
Dyslexic (D2)											-0.33	0.20
Slowed Down X Dyslexic											0.25	0.21
Amplified X Dyslexic											0.53	0.21
Both X Dyslexic											0.73	0.20
Random Part												
Level 2												
Intercept variance			0.11	0.03	0.36	0.08	1.79	0.38	1.81	0.39	1.83	0.38
Slope variance							0.42	0.09	0.42	0.10	0.42	0.09
Intercept-Slope covariance							0.12	0.03	0.12	0.03	0.12	0.03
Level 1												
Variance is Binomial with denominator = 1					$var(idL_{ij}\|\pi_{ij}$							
D	8360.07		8220.42		5335.96		4958.18		4920.15		4907.22	
D(ϕ)	8359.09		8167.94		5271.46		4835.14		4793.9		4777.51	
pD(D-D (ϕ))	0.98		52.47		65		123.04		126.24		129.71	
DIC	8361.06		8272.89		5400.45		5081.22		5046.39		5036.93	

labels as the target for the classification. At the same time, there was no significant main effect of group and the boundaries for the entire sample as predicted by M3 (see Table 4) deviated from the boundaries predicted by M4 for each group. To investigate the impact of these differences an additional QDA classification was performed using the predicted labels on the level of the sample.

The results for the sample are shown in Figure 8 and Table 6 that also includes the results for the predicted labels of M4 for each group of participants. What becomes apparent

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

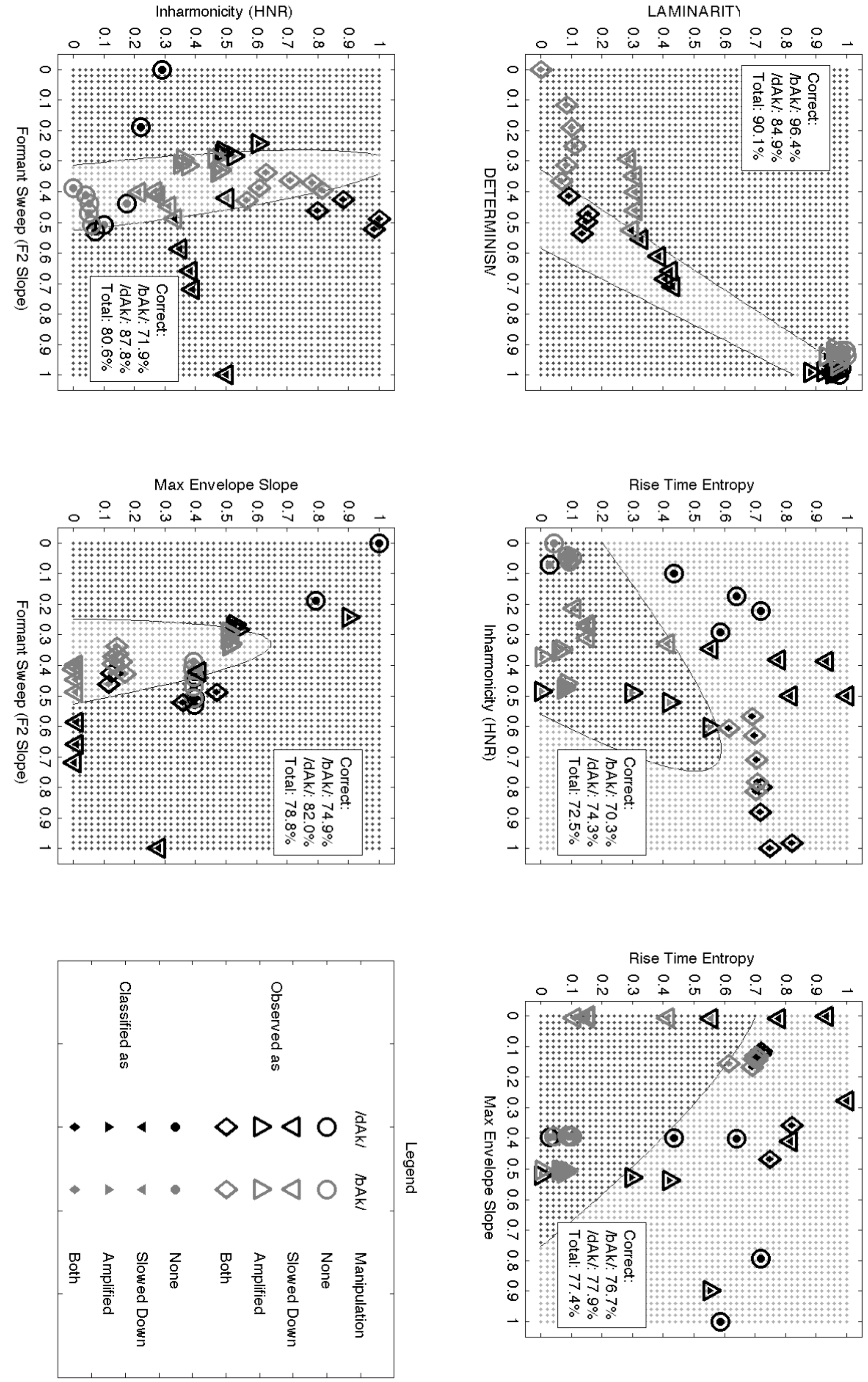

Figure 8. Results of the classification of the stimuli used in the experiment by Quadratic Discriminant Analysis. Targets of the classification were the labels predicted for the sample (M3, table 4). The panels show how the stimuli were observed (outer marker) and how they were categorised by QDA based on different pairs of measures (inner marker).

is that the Complexity measures outperform the other measures no matter which sequence of target labels is used.

Table 5

Predicted Odds (π) for Perceiving /dAk/ for all Participants from MCMC Model Estimation (Median of 150,000 iterations yielding 95% CI) for each stimulus and Acoustic Manipulation (M3 of Table 4). When the Lower CI Limit Exceeded 0.5 the Target for QDA was /dAk/, Otherwise it was /bAk/.

Stimulus	*Formant Onset (Hz)*			*Predicted Odds (π) for Perceiving /dAk/ per Manipulation*			
	F1	*F2*	*F3*	*None (CI*	*Slowed (CI*	*Amplified (CI*	*Both (CI*
/bAk/	440	1100	2700	0.06 (0.04, 0.08)	0.08 (0.05, 0.12)	0.13 (0.07, 0.24)	0.18 (0.08, 0.35)
2	|	1178	|	0.12 (0.08, 0.17)	0.15 (0.09, 0.25)	0.25 (0.13, 0.43)	0.32 (0.15, 0.57)
3	|	1255	|	0.23 (0.16, 0.33)	0.28 (0.17, 0.44)	0.43 (0.24, 0.64)	0.51 (0.27, 0.76)
4	|	1333	|	0.40 (0.28, 0.54)	0.47 (0.29, 0.66)	0.62 (0.38, 0.81)	0.70 (0.42, 0.88)
5	|	1411	|	0.60 (0.44, 0.74)	0.66 (0.45, 0.82)	*0.78 (0.56, 0.91)*	*0.84 (0.60, 0.95)*
6	|	1489	|	*0.77 (0.61, 0.87)*	*0.81 (0.63, 0.92)*	0.89 (0.72, 0.96)	0.92 (0.75, 0.98)
7	|	1567	|	0.88 (0.76, 0.94)	0.91 (0.77, 0.96)	0.95 (0.84, 0.98)	0.96 (0.86, 0.99)
8	|	1644	|	0.94 (0.86, 0.98)	0.95 (0.87, 0.98)	0.98 (0.91, 0.99)	0.98 (0.92, 1.00)
9	|	1722	|	0.97 (0.93, 0.99)	0.98 (0.93, 0.99)	0.99 (0.96, 1.00)	0.99 (0.96, 1.00)
/dAk/	440	1800	2700	0.99 (0.96, 1.00)	0.99 (0.97, 1.00)	0.99 (0.98, 1.00)	1.00 (0.98, 1.00)

[L] Lower CI limit ≥ 0.5 threshold (used as observed classification boundary)

[π] Predicted Median Odds ≥ 0.5 threshold

[U] Upper CI limit ≥ 0.5 threshold

9. Conclusion and Discussion

There are two clear and novel results to be discussed: 1. A difference between dyslexic and average readers in labelling some of the manipulated stimuli on the continuum is observed. 2. The Interaction-dominant measures outperform the other measures when used as features for a simple classifier that has as its targets several variations of plausible labels for each stimulus.

The first result entails the dyslexic readers identifying stimulus 4 as /dAk/ with 95% confidence above chance when the stimulus is either amplified or slowed down and subsequently amplified. It is thus not the case that dyslexic readers "benefit" from the manipulations in terms of their speech perception becoming more like that of average readers, instead, they perceive the boundary one continuum step earlier than average readers do whenever amplification is applied to the stimuli. It should be noted though that this "earlier" boundary perception is not the origin of the significant interaction effects between stimulus type and experimental group: the confidence intervals of the groups overlapped at

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Table 6

Quadratic Discriminant Analysis for Different stimulus Feature Combinations Based on Average Labelling by the Entire Sample, the Average Readers Group and the Dyslexic Readers Group. Numbers Represent Percentage Correctly Classified with CI

Group	Feature combination	Correct as /bAk/		Correct as /dAk/		Overall correct	
		Median	$CI_{.95}$	*Median*	$CI_{.95}$	*Median*	$CI_{.95}$
Sample	*LAM / DET*	96.4%	11.0%	85.0%	13.8%	90.1%	7.1%
	RFTe / HNR	70.7%	18.5%	72%	15.6%	72.6%	7.9%
	RFTe / mxENV Slope	76.7%	17.2%	77.9%	19.6%	77.3%	7.1%
	HNR / F2 Slope	72.4%	17.7%	87.3%	16.0%	80.6%	9.9%
	mxENV Slope / F2 Slope	75.0%	18.1%	82.0%	20.2%	78.9%	8.9%
Average readers	*LAM / DET*	96.2%	13.0%	88.2%	13.9%	91.4%	8.7%
	RFTe / HNR	75.6%	17.8%	72.1%	10.9%	73.5%	5.9%
	RFTe / mxENV Slope	80%	18.6%	81.1%	17.6%	82.2%	6.6%
	HNR / F2 Slope	67.6%	19.5%	86.2%	15.9%	78.7%	9.8%
	mxENV Slope / F2 Slope	69.9%	22.3%	75.1%	25.7%	73.0%	11.3%
Dyslexic readers	*LAM / DET*	96%	16.1%	87.6%	13.5%	90.0%	9.0%
	RFTe / HNR	76.7%	19.2%	72.1%	12.0%	73.7%	4%
	RFTe / mxENV Slope	85%	18.2%	81.1%	19.4%	82.3%	8.9%
	HNR / F2 Slope	63.9%	18.3%	86.7%	17%	78.7%	9.2%
	mxENV Slope / F2 Slope	73.6%	20.1%	78.6%	23.2%	76.9%	11.3%

these stimuli. Significant differences in odds for perceiving /dAk/ between the groups were observed for stimulus 6 (Amplified) and stimulus 5 (Both). This interaction is not likely to influence the actual labelling of the stimulus since both groups would label it /dAk/ above chance with 95% confidence. This difference would be noticed when the stimuli were presented to the same person many times in which case a dyslexic reader would label stimulus 5 (Both) about 9/10 times as /dAk/ and an average reader about 7/10 times. A similar result was found by Hasselman (2014a) where it was suggested that applying some manipulations may actually reduce the accuracy of identification and discrimination of stimuli because it biases perception towards /dAk/. The second result concerns the performance of a simple classifier (QDA) employed to label the stimuli as participants in the experiment using several different measures extracted from those stimuli. The classifier performed best when the Interaction Dominant dynamics measures (Determinism and Laminarity of the recurrence plot) were used. In fact, the classification of /bAk/ was almost perfect. Upon examination the misclassified stimulus was always the stimulus exactly on the

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

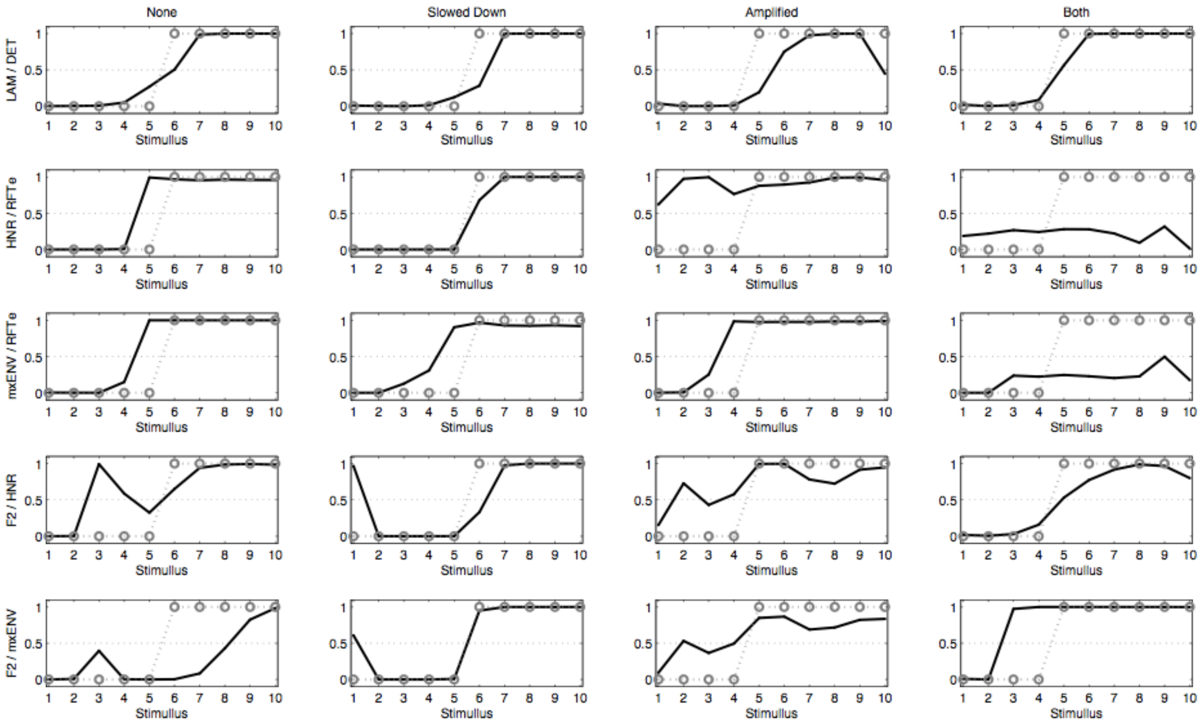

Figure 9. QDA class probabilities for each stimulus type and feature combination (black lines). The grey dotted lines are the targets for QDA based on the sample (Table 6).

boundary (i.e. with the $CI_{.95}$ neither above nor below 0.5). What is the implication of these findings for the two deficit hypotheses associated with the F2 Slope / HNR measures (auditory temporal processing deficit) and mxENV Slope / RFTe measures (rise time perception deficit)? First of all, these measures do yield different values that appear to differentiate the stimuli by rank on the continuum as well as by type of manipulation (see Figures 1, 2, 3 and Table 2). In other words they have the potential to be used for identification. In fact, the classification results, expressed as % correct are not disastrous when these measures are used and many stimuli are indeed labelled as human participants would. Some of these correct classifications may be expected from the way the stimuli are constructed. After all, this was done by manipulating the onset of the F2 while keeping everything else constant. The amplification of specific spectral features was also expected to have an effect on the amplitude envelope. Slowing down the signal obviously has an influence on the time it takes for the envelope to reach its maximal value. This makes the better classification result by measures based on an analysis of recurrences in a reconstructed phase space that hypothetically represents the dynamical system that produced the signal even more remarkable. Crucial is of course whether the classification approximates the labelling curve including the perceived boundary, the point at which participants switch from /bAk/ to /dAk/. Figure 9 shows the QDA class membership estimates for the targets based on the entire sample (Table 5). The grey dotted line represents the targets and the black line the QDA estimate for that stimulus type and feature combination. It is clear that the labelling curve is best approached by the estimates from the Interaction-dominant dynamics

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

measures (LAM / DET). The other curves at times show curves that are considerably off the mark.

The classification results presented here are comparable to those obtained by Little et al. (2007) when comparing RQA-derived measures with other features to classify speech recordings into healthy and disordered ones. In such a clinical context the benefit of roughly 10% more accurate detection of disordered speech is immediately apparent. In the present case one might object that none of the feature combinations provided a perfect match. Granted, it is not perfect. However, I do not argue that humans use a neurological equivalent of QDA to identify speech sounds. I want to show that it is very unlikely that they simply analyse (relative) frequency changes or amplitude envelopes and somehow match them to stored collections of frequencies and amplitude patterns. What I have shown here is that it is much more likely a higher order perception variable, similar to the higher order action control variables like Synergies and the Uncontrolled Manifold discussed earlier. The RQA measures provide such a higher order quantification of speech dynamics. The recurrence plot in a way is a holistic representation of all the temporal patterns, or correlations present in the speech signal. Contrary to the other derived measures, the values for DET and LAM remain constant over several continuum steps, changing slowly from stimulus to stimulus. Changes in DET occur sometimes around the perceived boundary. If we imagine for a moment recurrences in 1-D, so recurrences of actual recorded values, not 3-D coordinates, this might be the point where the F2 onset has increased by such an amount that it exceeds the threshold, or radius, so it is no longer registered as a recurrent point. The acoustic manipulations seem to have little influence on the rate with which the measures change, which is also understandable from the point of view of distances in phase space: The manipulations change (some) of the values in the signal relatively by equal amounts, which means the recurrences will occur in a different region of phase space, but they will recur nonetheless. Of course the effects of the manipulations on the labelling by participants were also small and concerned a shift of one step on the continuum.

This result does, however, not evidence against perception as storage and matching. It is still possible to argue that these higher order variables must be the information stored or represented somewhere in the brain waiting to be matched as we perceive them.

Information storage and retrieval: The similarity recognition problem

A popular application many people use on their smartphones called Shazam (Wang, 2003) is capable of analysing music being played in the environment and after a few seconds provide the name of the song and the artist who performed it. Interesting features are that it does not matter which part of the song is analysed and that as long as the recording being played exceeds background noise and is in the Shazam database, a few seconds of analysis are enough to yield almost 100% accuracy. The search and match time is reported to be between 5-500 milliseconds. Wang (2003) explains how the algorithm works. Based on a

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

sound recording a unique time-coded fingerprint is extracted from the spectrum and is stored in a database. When a song needs to be recognised, a smart search algorithm can quickly find likely candidates to which the small sample of the fingerprint belongs. The fingerprints are so unique the song is quickly identified. And this is exactly the reason why this is an unlikely model for speech perception. For instance: A song, of which the original studio recording is in the database, will not be recognised when a sample of a live recording by the same artist, not present in the database is searched for. Even a studio recording of the same song by the same artist, but with a different audio mix will not be recognised if it is not stored in the database. This is the equivalent of not being able to generalise a speaker's morning voice to his or her normal voice, in which case: "Hello darling" might become a confusing sentence when not aided by visual perception. Another reason to doubt speech perception works like this is that humans are generally not very good at accurately reconstructing a sentence when presented with just one or two words.

This problem of generalisation is one of many problems identified with the notion of perception as matching incoming information to stored information (see e.g. Chemero, 2009; Clark, 1999; Haselager, de Groot & van Rappard, 2003). Even if one wants to propose that we just store everything we hear from the day we are capable of doing so and disregard the fact that the amount of information to be stored would become infinitely large, it means we cannot understand someone the first time we meet him or her. We first have to store into a database the fingerprint of his or her utterances, using different speaking voices! Very unlikely indeed. This is called the problem of recognising similarity and Merleau-Ponty described it as follows: "*An impression can never by itself be associated with another impression. Nor has it the power to arouse others. It does so only provided that it is already understood in the light of the past experience in which it co-existed with those which we are concerned to arouse*." (Merleau-Ponty 1962, p, 1).

There is of course also technology that claims generalisation and promises to be able to recognise the natural speech of the user and follow spoken commands. One such example is the Siri personal assistant software that comes installed with the iPhone4s (Apple Inc., 2011) or Google Voice (Simonite, 2012). Generalisation in this and similar systems is achieved by using context: The meaning of a set of commands is narrowed down, if possible a-priori. There are just a limited number of ways to issue a command signifying you want to start to dictate an email. Moreover Siri uses syntactical structures to narrow down the options and conducts internet searches to figure out what it is you want it to do. If you are in New York and want directions to an address in Queens, the assistant uses that geographical knowledge and does not start to search for street addresses globally. In a way the degrees of freedom of the search are constrained by the context, but also by the physical features and purpose of the agent. The command uttered by the user is not just compared to a collection of stored commands in the device, nor are the agent's responses to the commands. Recognition and response are rather constructed from the meaning laden environment of the agent (geographical location, context of the interface, syntactic structure, a range of internet databases). At first glance this description seems quite similar to Merleau-Ponty's notion that

the representation of the environment is an unnecessary assumption in understanding intelligent behaviour. When one examines how human perception and action is constrained by the physical features of the body and environment of the agent, representation turns into an odd concept (Dreyfus, 2002). Even so, Siri does not generalise enough, because users who speak with accents and dialects have a hard time commanding the assistant (Effron, 2011).

Let's consider how this constraining of degrees of freedom might work for human speech perception by taking seriously a remark by Stetson: "*Speech is rather a set of movements made audible than a set of sounds produced by movements*" (1951, p. 33). This implies the focus on auditory signal analysis as a model for speech perception noted in the introduction is indeed unwarranted. One should at least consider the complex system that produces the signal when theorising about perception of the signal. There is evidence that a close bi-directional perception-action coupling exists when speech perception and production are concerned. In a series of experiments Perkell et al. (2004a; 2004b) have shown that the distinctness, or quality of a produced vowel contrast by a speaker, is related to the quality of the perception of that contrast by the same speaker. In other words, speech production will constrain speech perception and vice versa. Some of these notions have been incorporated in the successful DIVA (Directions Into Velocities of Articulators) model of speech production (Guenther & Perkell, 2004). In short, this model learns to produce speech by tuning, or constraining its motor output to auditory targets it is presented with.

This brings us back to the discussion of the classification performance of the measures obtained from the speech signal by RQA and their interpretation by virtue of Takens' theorem. These measures represent the dynamical behaviour of the system that produced the signal we used to reconstruct phase space (by topological equivalence). In this example the system in question is of course the speech apparatus that controls (turbulent) airflow through bodily cavities. This is the same interpretation used by Little et al. (2007) who hypothesised the dynamics of the speech signal would carry information about a speech apparatus that is not functioning optimally due to, for instance, a tumour obstructing airflow or an affliction of the vocal cords. In a way, the phase space dynamics operationalised by the recurrence matrix represent the interaction between speech gesture and airflow that results in the propagating pressure wave that is the subject of analysis. The measurement level is the boundary between action and perception at a global, or holistic level. This claim is quite modest when compared to Stephen et al. (2009) who showed the reconstructed phase space dynamics of finger movements are associated to changes in the cognitive system induced by insight in problem solving. In the present study the most noticeable difference between producing /bAk/ and /dAk/ is in the alveolar place of articulation when producing /dAk/ (tongue moves from the front of the palate downwards), whereas /bAk/ is pronounced without tongue movements. This difference in speech gesture is generally thought to be associated to the slope of the second formant (F2) in the spectrogram, or at least the relative changes in the first three formants in the speech signal. In reality the place of articulation is not sufficient to describe the gesture which is clearly much more complex. Consider the difference between

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

pronouncing the sounds /dA/ and /lA/ or /bA/ and /wA/. Lip movements, tongue shape, timing of pressure release, relative timing of these actions, they are all important. Recent studies in speech signal analysis have shown the frequency domain obtained by Fourier decomposition may not be the information used by the neural systems of mammals to perceive sounds, whereas the Hilbert decomposition in slow varying envelope and fast varying fine time structure (the analytic signal), may be the more likely candidate (Smith, Delgutte & Oxenham, 2002). The Rise-Time Perception Deficit Hypothesis of dyslexia (cf. Goswami et al., 2002) is partially based on these findings. The fact that the speech signal is the product of the fast analytic signal and the slow changing envelope is not considered by the theory. In any case, the idea of speech sounds being stored in memory as strings of abstract phonetic units like formants has come under attack (see Port, 2007 for a review of arguments against this notion). The results of the present study add to the existing critiques the notion that higher-order variables need to be considered.

It is of course important to replicate these findings with other stimuli and other samples of participants. Interestingly, the analysis presented here can be performed post hoc on any speech identification study already published. The measures can be extracted from any signal and the QDA can be applied using the observed labels found in the study as targets for the classification.

References

Assmann, B., Romano, M. C., Thiel, M., & Niemitz, C. (2007). Hierarchical organization of a reference system in newborn spontaneous movements. *Infant behavior & development*, *30*(4), 568-86. doi:10.1016/j.infbeh.2007.0004

Apple, Inc. (2011). Siri. Your wish is its command. Retrieved from http://www.apple.com/iphone/features/siri.html

Bianciardi, M., Sirabella, P., Hagberg, G. E., Giuliani, A., Zbilut, J. P., & Colosimo, A. (2007). Model-free analysis of brain fMRI data by recurrence quantification. *NeuroImage*, *37*(2), 489-503. doi:10.1016/j.neuroimage.2007.05.025

Boets, B., Ghesquière, P., van Wieringen, A., & Wouters, J. (2007). Speech perception in preschoolers at family risk for dyslexia: relations with low-level auditory processing and phonological ability. *Brain and language*, *101*(1), 19-30. doi:10.1016/j.bandl.2006.06.009

Breznitz, Z. (2003). Speed of processing of the visual–orthographic and auditory–phonological systems in adult dyslexics: The contribution of “asynchrony” to word recognition deficits. *Brain and Language*, *85*(3), 486-502. doi:10.1016/S0093-934X(03)00071-3

Browne, W.J. (2009) MCMC Estimation in MLwiN v2.1. Centre for Multilevel Modelling, University of Bristol.

Chemero, A., & Turvey, M. T. (2007). Complexity, hypersets, and the ecological perspective on perception-action. *Biological Theory*, *2*(1), 23–36. MIT Press. Retrieved from http://www.mitpressjournals.org/doi/abs/10.1162/biot.2007.2.1.23

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3(9)*, 345-351.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: tapping to the beat. *Cortex; a journal devoted to the study of the nervous system and behavior*, *45*(1), 119-30. Elsevier Srl. doi:10.1016/j.cortex.2007.09.008

Corriveau, K. H., Pasquini, E. S., & Goswami, U. (2007). Basic Auditory Processing Skills and Specific Language Impairment: A New Look at an Old Hypothesis. *Journal of Speech, Language, and Hearing Research*, *50*(3), 647. ASHA. Retrieved from http://jslhr.highwire.org/cgi/content/abstract/50/3/647

Dale, R., & Spivey, M. J. (2006). Unraveling the Dyad: Using Recurrence Analysis to Explore Patterns of Syntactic Coordination Between Children and Caregivers in Conversation. *Language Learning*, *56*(3), 391-430. doi:10.1111/j.1467-9922.2006.00372.x

De Graag, J. A., Cox, R. F. A., Hasselman, F., Jansen, J., & De Weerth, C. (n.d.). Functioning within a relationship: Mother-Infant Synchrony and Infant Sleep. *Infant Behavior & Development*.

Diniz, A., Wijnants, M. L., Torre, K., Barreiros, J., Crato, N., Bosman, A. M. T., Hasselman, F., et al. (2011). Contemporary theories of 1/f noise in motor control. *Human movement science*, *30*(5), 889-905. Elsevier B.V. doi:10.1016/j.humov.2010.07.006

Dreyfus, H.L. (2002). Intelligence without representation – Merleau-Ponty's critique of mental representation The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences, 1*, 367-383. doi: 10.1023/A:1021351606209

Effron, L. (2011). Siri Lost in Translation with Heavy Accents. *ABC News*. Published 2011-10-28. Retrieved 2012-01-06 from: http://abcnews.go.com/Technology/siri-lost-translation-heavy-accents/story?id=14834111#.Txgs9JhCjq-

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, *2*(4), 460–493. Springer.

Feldman, M. (2008). Theoretical analysis and comparison of the Hilbert transform decomposition methods. Mechanical Systems and Signal Processing, 22(3), 509–519. doi:10.1016/j.ymssp.2007.09.013

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*(3), 361. Psychonomic Society Publications. doi:10.2964/jsik.kuni0223

Gibson, J. J. (1979). *The ecological approach to visual perception* (p. xiv, 332 pp.). Boston, MA: Houghton, Mifflin and Company; US.

Goswami, U. (2006). Sensorimotor impairments in dyslexia: getting the beat. *Developmental Science*, *9*(3), 257–259. Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-7687.2006.00485.x/abstract

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in cognitive sciences*, *15*(1), 3-10. Elsevier Ltd. doi:10.1016/j.tics.2010.10.001

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szűcs, D. (2010). Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. *Developmental Science*, *1*, 34-43. doi:10.1111/j.1467-7687.2010.00955.x

Goswami, U., Thomson, J., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., & Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National*

Academy of Sciences of the United States of America, *99*(16), 10911-6. doi:10.1073/pnas.122368599

Guenther, F. H., & Perkell, J. S. (2004). A neural model of speech production and supporting experiments. *From Sound to Sense' Conference, 98-106*. doi=10.1.1.121664

Guiard, Y. (1993). On Fitts's and Hooke's laws: Simple harmonic movement in upper-limb cyclical aiming. *Acta Psychologica*, *82*(1-3), 139-159. doi:10.1016/0001-6918(93)90009-G

Haselager, P., de Groot, A., & van Rappard, H. (2003). Representationalism vs. anti-representationalism: A debate for the sake of appearance. *Philosophical Psychology, 16(1)*, 5-2 doi: 10.1080/0951508032000067761

Hasselman, F. (2014a). Context Relativity of Impaired Performance (Beyond The Boundary - Chapter 2). fig**share**. http://dx.doi.org/10.6084/m9.figshare.928517

Hasselman, F., (2014b). Beyond The Boundary - 4th Chapter: Supplemental Materials. Retrieved from Open Science Framework, osf.io/2r5eh

Hasselman, F. (2014c). Context Relativity of Impaired Performance (Beyond The Boundary - Chapter 5). fig**share**. http://dx.doi.org/10.6084/m9.figshare.928517

Huss, M., Verney, J., Fosker, T., Mead, N., & Goswami, U. (2010). Music, Rhythm, Rise Time Perception and Developmental Dyslexia: Perception of Musical Meter predicts Reading and Phonology. *Cortex*, 1-16. Elsevier Srl. doi:10.1016/j.cortex.2010.07.010

Lamb, H. (2004). *The dynamical theory of sound*. Dover Pubns.

Lichtwarck-Aschoff, A., Hasselman, F., Cox, R. F. A., & Granic, I. (n.d.). The Unsolid Foundation of Successful Intervention: Destabilization of Parent Child Interactions as a Necessary Phase in Effective Treatment. *Nonlinear Dynamics in Psychology and Life Sciences*.

Little, M. A., McSharry, P. E., Moroz, I. M., & Roberts, S. J. (2006). Nonlinear, Biophysically-Informed Speech Pathology Detection. *Proceedings of IEEE ICASSP 2006*. IEEE Publishers: Toulouse, France.

Little, M. A, McSharry, P. E., Roberts, S. J., Costello, D. a E., & Moroz, I. M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical engineering online*, *6*, 23. doi:10.1186/1475-925X-6-23

Livingstone, M. S., Rosen, G. D., Drislane, F. W., & Galaburda, A. M. (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(18), 7943-7.

Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, *438*(5-6), 237-329. doi:10.1016/j.physrep.2006.11.001

Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London and New York: Routledge and Kegan Paul.

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Prentice-Hall Englewood Cliffs, NJ.

Moin, P., & Kim, J. (1997). Tackling turbulence with supercomputers. *Scientific American*, *276*(1), 46–52.

Pasquini, E. S., Corriveau, K. H., & Goswami, U. (2007). Auditory processing of amplitude envelope rise time in adults diagnosed with developmental dyslexia. *Scientific Studies of Reading*, *11*(3), 259–286.

Perkell, J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., et al. (2004a). The distinctness of speakers' /s/-/S/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language and Hearing Research* ,*47*, 1259-1269.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L. Stockmann, E., Tiede, M., et al. (2004b). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *Journal of the Acoustical Society of America, 116*, 2338-234

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25(2)*, 143-170. doi:10.1016/j.newideapsych.2007.02.001

Porter, R. J., & Hogue, D. M. (1998). Nonlinear dynamical systems in speech perception and production. *Nonlinear Dynamics, Psychology, and Life Sciences*, *2*(2), 95–131.

Rasbash, J., Charlton, C., Browne, W.J., Healy, M. and Cameron, B. (2009) MLwiN Version 2.1. Centre for Multilevel Modelling, University of Bristol.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. *Psychological science: a journal of the American Psychological Society / APS*, *18*(5), 407-13. doi:10.1111/j.1467-9280.2007.0191x

Riley, M. A., & Van Orden, G. C. (2005). *Tutorials in contemporary nonlinear methods for the behavioral sciences*. National Science Foundation.

Schinkel, S., Dimigen, O., & Marwan, N. (2008). Selection of recurrence threshold for signal detection. *The European Physical Journal Special Topics*, *164*(1), 45-53. doi:10.1140/epjst/e2008-00833-5

Schinkel, S., Marwan, N., & Kurths, J. (2007). Order patterns recurrence plots in the analysis of ERP data. *Cognitive neurodynamics*, *1*(4), 317-25. doi:10.1007/s11571-007-9023-z

Schinkel, S., Marwan, N., & Kurths, J. (2009). Brain signal analysis based on recurrences. *Journal of physiology, Paris*, *103*(6), 315-23. Elsevier Ltd. doi:10.1016/j.jphysparis.2009.05.007

Schinkel, S., Marwan, N., Dimigen, O., & Kurths, J. (2009). Confidence bounds of recurrence-based complexity measures. *Physics Letters A, 373(26)*, 2245-2250. doi:10.1016/j.physleta.2009.0045

Simonite, T. (2012). Google Puts Its Virtual Brain Technology to Work. A powerful new approach to artificial intelligence is ready to improve many Google products. *MIT Technology Review*. Published 2012-10-5. Retrieved from: http://www.technologyreview.com/news/429442/google-puts-its-virtual-brain-technology-to-work/

Scholz, J. P., & Schöner, G. (1999). The uncontrolled manifold concept: identifying control variables for a functional task. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, *126*(3), 289-306. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10382616

Schöner, G., & Kelso, J. (1988). Dynamic Pattern Generation in Behavioral and Neural Systems. *Science*, *239*(4847), 1513–1520.

Shockley, K., Baker, A. a, Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of experimental psychology. Human perception and performance*, *33*(1), 201-8. doi:10.1037/0096-1523.33.1.201

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326-332. doi:10.1037/0096-1523.29.2.326

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416(6876)*, 87-90. doi:10.1038/416087a

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.341v1 | CC-BY 4.0 Open Access | received: 31 Mar 2014, published: 31 Mar 2014

Stein, J. F. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia (Chichester, England)*, *7*(1), 12-36. doi:10.1002/dys.186

Stein, J. F., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in neurosciences*, *20*(4), 147–152. doi:10.1016/S0166-2236(96)01005-3

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, Action and Cognition. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1811–1832.

Stetson, R.H. (1951). *Motor phonetics: A study of speech movements in action* (2nd ed.). Amsterdam, North-Holland: Publication Company for Oberlin College.

Takens, F. (1981). Detecting strange attractors in turbulence. *Dynamical systems and turbulence, Warwick 1980*, 366–381. Retrieved from http://www.springerlink.com/index/B254X77553874745.pdf

Tallal, P. (1976). Rapid auditory processing in normal and disordered language development. *Journal of speech and hearing research*, *19*(3), 561-71. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/979217

Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature reviews. Neuroscience*, *5*(9), 721-8. doi:10.1038/nrn1499

Tallal, P., & Piercy, M. (1974). Developmental aphasia: the perception of brief vowels and extended stop consonants. *Neuropsychologia*, *12*(1), 69-7 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4821193

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological Basis of Speech: A Case for the Preeminence of Temporal Processing. *Annals of the New York academy of sciences*, *682*(1), 27–47.

Temple, E., Poldrack, R. a, Protopapas, A., Nagarajan, S., Salz, T., Tallal, P., Merzenich, M. M., et al. (2000). Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(25), 13907-12. doi: 10.1073/pnas.240461697

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human movement science*, *26*(4), 657-97. doi:10.1016/j.humov.2007.0002

Wang, A. (2003). An industrial strength audio search algorithm. International Conference on Music Information Retrieval (ISMIR) (Vol. 2). Retrieved from http://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

Webber Jr., C. L., & Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. *Tutorials in contemporary nonlinear methods for the behavioral sciences*, 26–9 Retrieved from http://www.nsf.gov/sbe/bcs/pac/nmbs/chap2.pdf

Webber Jr., C. L., Marwan, N., Facchini, A., & Giuliani, A. (2009). Simpler methods do it better: Success of Recurrence Quantification Analysis as a general purpose data analysis tool. *Physics Letters A*, *373*(41), 3753-3756. doi:10.1016/j.physleta.2009.08.052

Wijnants, M. L., Bosman, A. M. T., Cox, R. F. A., Hasselman, F., & Van Orden, G. C. (2011). Nested Timescales of Motor Control: a Trade-Off Study. In E. Charles & L. J. Smart (Eds.), *Studies in perception and action XI*. Taylor & Francis Group, LLC.

Wijnants, M. L., Bosman, A. M. T., Hasselman, F., Cox, R. F. A., & Van Orden, G. C. (2009). 1/F Scaling in Movement Time Changes With Practice in Precision Aiming. *Nonlinear dynamics, psychology, and life sciences*, *13*(1), 79-98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19061546

Zbilut, J. P., Giuliani, A., & Webber Jr., C. L. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters A*, *246*(1-2), 122-128. doi:10.1016/S0375-9601(98)00457-5

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3-29. doi:10.1037/0033-2909.131.1.3

van Beinum, F. J., Schwippert, C. E., Been, P. H., van Leeuwen, T. H., & Kuijpers, C. T. L. (2005). Development and application of a /bAk/–/dAk/ continuum for testing auditory perception within the Dutch longitudinal dyslexia study. *Speech Communication*, *47*(1-2), 124-142. doi:10.1016/j.specom.2005.0003