Chapter 2

Quantifying Movement Variability

2.1 Introduction

It has been stated in Chapter 1 that movement variability can be modelled and quantified using nonlinear tools mainly because the structures of the human physiology (lungs, neurons, etc.) suggest that many of their dynamics are controlled by nonlinear dynamics (Goldberger et al., 1990) and data from human movement is essentially chaotic ~~deterministic~~, meaning that it is neither deterministic nor stochastic (Hatze, 1986; Preatoni et al., 2010, 2013; Stergiou et al., 2006). Additionally, data from the human body is ~~generally~~ noisy, deterministic, stochastic or nonstationary (Newell and Slifkin, 1998). Therefore, in this chapter fundamentals of time series, nonlinear tools and nonlinear tools with real-data will be reviewed.

2.2 Fundamentals of time-series analysis

Biosignals from living systems can typically be nonstationary, nonlinear, deterministic chaotic and noisy (Caballero et al., 2014; Gómez-García et al., 2014; Harbourne and Stergiou, 2009; Hatze, 1986; Klonowski, 2007; Newell and Slifkin, 1998; Stergiou and

Decker, 2011; Stergiou et al., 2006; Wijnants et al., 2009). Therefore, it is important to provide fundamental definitions of time series which will be used thought the thesis.

2.2.1 Linear and nonlinear systems

Linear systems are proportional or additive. For example, the interaction between variables of a linear system are negligible whereas for a nonlinear system such interaction of variables can produce emergent properties ~~due to~~ arising from the initial conditions of the system (Klonowski, 2007).

2.2.2 Stationary and nonstationary signals

Stationary signals have the same mean and variance as time progress (e.g. a sinusoidal signal), however such stationary signal can also be changeable (e.g. alternative sinusoidal signal). In contrast, when statistics of the time series change with time then such a signal is known as nonstationary signal. Nonstationary signals are therefore characterised by ~~having~~ transients and drift~~s~~ over time. Examples of nonstationary signals are the time series of seasonal trends and changes (Kitagawa and Gersch, 1984), or Electroencephalography (EEG) signals which present different and changeable intensity over time (Klonowski, 2007).

2.2.3 Deterministic and stochastic systems

A deterministic system~~s means that~~ is predictable. Deterministic systems have ~~the characterising to have~~ a small number of variables of importance ~~in the system~~. Deterministic systems are modelled with linear ordinary differential equations and their initial conditions and constants. In contrast, stochatic systems are nonpredictable and therefore have ~~bigger number of~~ more variables of equal importance and ~~stochastic systems~~ are modelled with probability theory (Klonowski, 2007).

2.2.4 Deterministic chaotic time series

EXPAND

Deterministic signals can dramatically change with a slight change of initial conditions and then after a long time-scale, the signal can appear to be stochastic (Amato, 1992). Similarly, Klonowski 2007, p. 11 pointed out that "chaotic systems behave like they were stochastic but they are also deterministic", meaning that chaotic systems are predictable for a short time-scale but nonpredictable in a long time-scale because of the initial conditions of the systems. Then, Preatoni et al. 2013, p. 78, in experiments in sport science, mentioned that "variability is likely to have both deterministic and a stochastic origin". Therefore, it can be concluded that time series for human body movement are neither independent nor stochastic but deterministic chaotic (Harbourne and Stergiou, 2009; Stergiou and Decker, 2011; Stergiou et al., 2006).

[Is this right? Aren't they predictable over long time and not over short time?]

Why include 'deterministic' here?

2.3 Quantifying Movement Variability with Nonlinear Dynamics

2.3.1 Introduction

Previous studies have shown that movement variability is not considered as a undesired factor that creates errors but a signature for assessment of healthiness (associated with unhealthy pathological states) or skillfulness (associated with the functionality of movement) (Stergiou and Decker, 2011). Fundamentally, movement variability can be either quantified based on magnitude of the variability or the dynamics and complexity of the variability (Caballero et al., 2014). However, finding the right tools to quantify movement variability is still an open problem.

For instance, Preatoni et al. (2010, 2013) pointed out that conventional statistics (e.g. standard deviation, coefficient of variation, intra-class correlation coefficient) only quantify the overall variability. Also, Stergiou and Decker (2011) stated that statistical tools (e.g. mean, standard deviation and ~~the~~ range) are a measure of centrality, meaning such metrics are compared around a central point. Similarly, Coffey et al. (2011) pointed out that the use of means and standard deviations led to reduction of data and information is therefore discarded.

Additionally, one can apply frequency-domain tools to quantify movement variability. For example, Hatze (1986) proposed a measure of dispersion to quantify the deviation of motion from a certain reference using the Fourier series. However, deviations of motion are from angular coordinates (radians) and linear coordinates (meters) which made them an unacceptable fusion of variables. Vaillancourt et al. (2001) pointed out that it is rare for frequency and amplitude to differ in postural tremor of patients with Parkinson's disease but differences in time-dependent structures are apparent, and associated with a change of regularity of postural tremor. ~~Then,~~ Klonowski (2002, 2007, 2009) stated that frequency-domain tools require ~~to have~~ stationary data, otherwise using other type of data might create misleading results.

Therefore, applying either statistical tools or frequency-domain tools to quantify movement variability might create misleading results, specially when dealing with signals that are deterministic chaotic (Amato, 1992; Dingwell and Cusumano, 2000; Dingwell and Kang, 2007; Miller et al., 2006), considering that the subtle changes in the neuro-muscular-skeletal system are caused by influences of environmental changes, training ~~procedures~~ or latent pathologies (Preatoni et al., 2010, 2013) and that movement variability involves evolution of human movement and the exploratory nature of movement (Caballero et al., 2014; Stergiou and Decker, 2011). Hence, Caballero et al. (2014); Preatoni et al. (2010); Stergiou and Decker (2011) highlighted that movement variability can be better described and quantified with different nonlinear dynamics tools such as: largest Lyapunov exponent (Bruijn et al., 2009; Donker et al., 2007; Kurz et al., 2010; Yang and Wu, 2011), fractal analysis (Delignlères et al., 2003), entropy rate (Cavanaugh et al., 2010), Sample Entropy (SampEn) (Donker et al., 2007; Liao et al., 2008; Richman and Moorman, 2000; Stins et al., 2009; Vaillancourt et al., 2004), Approximate Entropy (ApEn) (Cavanaugh et al., 2010; Kurz and Hou, 2010; Pincus, 1991; Sosnoff et al., 2006; Sosnoff and Voudrie, 2009), Fuzzy Entropy (FuzzyEn) (Chen et al., 2007), Multiscale Entropy (MSE) (Costa et al., 2002), Permutation Entropy (PE) (Bandt and Pompe, 2002; Vakharia et al., 2015), Quadratic Sample Entropy (QSampEn) (Lake and Moorman, 2011), Amplitude-aware permutation entropy (AAPE) (Azami and Escudero, 2016), Detrended Fluctuation Analysis (DFA) (Gates and Dingwell, 2007, 2008; Hausdorff, 2009) and Recurrence Quantification Analysis (RQA) (Marwan, 2008; Trulla et al., 1996; Zbilut and Webber, 1992).

Having so ~~got~~ many nonlinear tools to measure movement variability (MV) led ~~made~~ Caballero et al. 2014, p. 67 to raise the following question: "Is there a best tool to measure variability?" which leads us to ~~formulate a further set of questions for this thesis on~~ ask what to measure in MV?, how to measure MV? and which tools are appropriate to measure MV?

2.3.2 What to measure in MV?

Vaillancourt and Newell (2002, 2003) stated that there is no universal increase or decrease in complexity for MV as a function of age or disease but a dependency with the task dynamics. For example, in a constant-force task (where the task dynamics is of low dimension), ~~generally~~ older adults present less complexity due their inability to introduce additional degrees of freedom in the neuromuscular system. However, there is an increase of complexity in older adults or unhealthy adults when the task dynamic is oscillatory because these type of adults have more difficulty to reduce the dimension output to a lower dimension which are the intrinsic dynamics of their resting state.

In contrast, inspired by Tononi et al. (1998) who modelled complexity in neural networks considering complexity versus regularity variables, Stergiou et al. (2006) proposed a model of complexity versus predictability ~~variables~~ for optimal human movement variability. The model of Stergiou et al. (2006) stated that higher complex movements are associated with rich behaviour of movements while lower complex movements are associated with poor behaviours of movements being too rigid or too unstable. Hence, higher complex ity movements are ~~therefore~~ characterised by chaotic systems, while lesser complexity of movement~~s~~ is ~~are~~ characterised either as noisy systems or periodic systems (having either low amounts of predictability or hight amounts of predictability) (Stergiou et al., 2006). [So... what is the answer to the 'what to measure' question?]

2.3.3 Which nonlinear tools are appropriate to measure MV?

Considering the model of Stergiou et al. (2006) for movement variability, where complexity and predictability variables of a system can characterise and quantify movement variability, it is important to find, to understand and to apply the right tools that measure such variables.

Originally, Pincus (1995, 1991) proposed Approximate Entropy (ApEn) to quantify regularity of time series. Then, Richman and Moorman (2000) due to self-matching found that the algorithm of ApEn could evoke ~~match itself to avoid~~ the occurrence of ln(0) which made ApEn dependant on the available data. ~~for which~~ Sample Entropy (SampEn) was ~~were~~ proposed as an algorithm that does not consider self-matching. Hence, SampEn values are independent of the length of time series and its algorithm is simpler than ApEn. Then, instead of using single statistics, Costa et al. (2002) proposed Multiscale Entropy (MSE) ~~algorithm~~ which computes SampEn of consecutive coarse-grained time series of the original time series defined by the scale factor, τ. With MSE algorithm, (Costa et al., 2002) noted that pathology dynamics for time series of heartbeat intervals are associated with reduction of complexity. Therefore, Costa et al. 2002, p. 3 concluded that physiologic al complexity is associated with ~~to~~ the adaptive capacity of the organism, diseases states and aging which "may be defined by a sustained breakdown of long-term correlations and loss of information". Essentially, entropy measures (AppEn and SampEn), quantify regularity and complexity of time series (Preatoni et al., 2013). However, Goldberger (1996) mentioned that the increase of irregularity in time series is not ~~a~~ synonymous with ~~of~~ increase of physiologic al complexity. Similarly, an increase of ApEn or SampEn, "implying increase of irregularity and decrease in predictability" (Goldberger et al., 2002, p. 25), is not synonymous with ~~of~~ an increase of dynamical complexity when analysing physiology signals (Costa et al., 2002). Hence, Goldberger et al. (2002) demonstrated that ApEn as a regularity statistic is not a direct index of physiologic al complexity where, for example, a randomised time series of an healthy heartbeat with multi-scale and complex patterns of variability show a higher value of ApEn even though ~~being that~~ the time series is less complex. Therefore, Goldberger et al. 2002, p. 24 concluded that the loss of physiologic al complexity can be "better assessed using other measures which can detect and quantify the presence of long-range correlations in nonstatiorany series." Hence, Costa et al. (2002); Goldberger et al. (2002); Vaillancourt and Newell (2002) concluded that ApEn and SampEn do not necessary show the right representation of what they ~~are~~ intend ed to measure.

Therefore, considering the previous cons of ApEn, SampEn and MSE, Detrended Fluctuation Analysis (DFA), which is based on analysing fractal features, can quantify long-term correlations of time series (Peng et al., 1995). DFA is calculated as the root mean square fluctuation of an integrated and detrended time series and it is represented by a scaling exponent, α, which is an indicator for roughness of time series, e.g. "the larger the value of α, the smoother the time series (Peng et al., 1995, p. 83). However, ~~only using~~ DFA can result in a false conclusion for long term correlations in the time series (Rangarajan and Ding, 2000, p. 5001), therefore DFA "can falsely classify certain type of time series as fractals" (Wijnants et al., 2009, p. 80). With that in mind, Wijnants et al. (2009) proposed the use of RQA as a technique that does not present any constraints with regards to length size, stationary or statistical distribution of the time series. ~~Nonetheless,~~ Wijnants et al. (2009) highlighted that SampEn index is calculated over the sequential values of the time series, whereas Shannon entropy in RQA which is computed over the distribution of deterministic lines in the Recurrence Plots (RP) (Marwan, 2008; Trulla et al., 1996; Zbilut and Webber, 1992). Similarly, Rhea et al. (2011) highlighted that algorithms to compute entropy measures are different since ApEn and SampEn are approximations of the Kolmogorov-Sinai Entropy computing the likelihood that a template pattern repeats in the time series while RQAEn is derived from Shannon entropy and is computing number of line segments of varying length in the RP. Even ~~though~~ with those differences in the algorithms, smaller values of recurrence percentage of the RQA shown the increase of practice of movement dynamics, concluding that such recurrence percentage can ~~is~~ indicate~~or of~~ increase in ~~of~~ system stability (Wijnants et al., 2009).

Another tool to measure variability is the largest Lyapunov exponent (LyE) which is used to "quantify the exponential separation of nearby trajectories in the reconstructed state space of a time series" (Stergiou, 2004, p. ??). For instance, "LyE from a stable system with little to no divergence will be zero (e.g. sine wave)" and "LyE for an unstable system that has highest amount of divergence will be positive and relative hight in value (e.g. 0.469 for random noise)" and for chaotic systems like the Lorenz system, LyE is in between the two of the previous extremes (LyE$\approx$ 0.1) (Miller et al., 2006, p. 2874). However, LyE requires to be validated using surrogation (Dingwell and Cusumano, 2000; Miller et al., 2006).

Measuring human movement variability requires a combination of the pros and cons of many of the previous tools that analys~~is~~e either (i) the dynamic complexity or (ii) the degree of regularly, stability or predicability in a system (Goldberger et al., 2002; Harbourne and Stergiou, 2009; Stergiou and Decker, 2011). For instance, Rangarajan and Ding (2000) stated the use of both spectral analysis and random walk analysis, the base of DFA, is a better approach than only using one tool which can lead to false conclusion for long term correlations in the time series. Similarly, Wijnants et al. (2009) selected different tools (spectral analysis, standard dispersion analysis, DFA, RQA and SampEn) to quantify movement variability that can complement the strengths of some of them and also compensate the weakness of others. Recently, Caballero et al. (2014) proposed the unification of different tools to address every aspect of the dynamics of a systems and the characterisation of the variability.

Although, there is no best tool to measure movement variability and an unification of tools to quantify human movement variability is still an open question, finding the right tool to measure movement variability for an specific problem, and knowing its strengths and weakness of such tool is one of the research questions for this thesis.

2.4 Nonlinear analyses with real-world data

Recently, Huffaker et al. (2017) ~~only highlighted~~ pointed out that one of the caveats when applying nonlinear time series analysis tools is its unreliability when the estimated metrics come from real-world data which is generally short, noisy and nonstationary. Similarly, Preatoni et al. (2013) mentioned the limitations of ~~the use of~~ nonlinear analyses in sport activities where data required to be large (e.g. number of trials, duration of the experiment and sampling frequency). ~~Whereas~~ Caballero et al. (2014)~~, providing further a investigation,~~ ~~stated the~~ argued that there are weaknesses of different nonlinear tools regarding the characteristics of the time series such as nonstationarity, length data size, nois~~y~~e sampling rate. However, in the work of Huffaker et al. (2017), Preatoni et al. (2013) and Caballero et al. (2014) no further exploration of the metrics with real-world data is presented.

2.4.1 Nonsationarity

Nonstationarity of time series signals might create spurious increase or decrease in the metrics of nonlinear tools. For instance, Costa et al. (2007) noted that ~~nonsatiornary~~ nonstationarity in the signals might alter the increase of irregularity of signals for the shortest scales when applying MSE. Also, Dingwell and Cusumano (2000) reported nonstationar~~y~~ity in time series when using LyE, which required to be validated using surrogation to ensure the robustness of the metric. Hence, Caballero et al. (2014) reported three options when dealing with nonstaionary data: (i) remove nonsatiornary data, (ii) use empirical mode decomposition (EMD), and (iii) apply nonlinear tools, such as DFT and RQA, which are less sensitive to nonstationary data.

To remove nonstationary data, for example, Carroll and Freedman (1993) suggested to remove the trends or to eliminate the ~~first 20 seconds of~~ initial data samples to ignore the trend of time series. Hence, van Dieën et al. (2010), in experiments with center of pressure movements in seated balancing, discarded the first 5 seconds of the time series in the start of the measurement ~~to avoid nonstationary of the data~~.

Also, nonstatioinary time series can be treated with Empirical Mode Decomposition (EMD) method which decompose nonlinear, nonstatioanry signals into their intrinsic frequency components (Huang et al., 1998; Wu and Huang, 2004, 2009). Hence, Costa et al. (2007); Flandrin et al. (2004) tested ~~that~~ whether EMD is a robust method for detrending and denoising time series and ~~highlighted~~ noted that EMD does not require selection of input parameters. However, the reliability of EMD methods is still an open problem. For instance, an extension of EMD called Multivariate Empirical Mode Decomposition

(MEMD) has been proposed to analyse multiple time series (Mandic et al., 2013; Rehman and Mandic, 2010). See (Bonnet et al., 2014; Costa et al., 2007; Daubechies et al., 2011; Mert and Akan, 2018; Wu and Hu, 2006) ~~for applications of EMD~~.

Finally, one can use ~~of~~ nonlinear tools that are unaffected by ~~nonsttarionary~~ nonstationarity of time series such as Detrended Fluctuation Analys (DFA) (Hausdorff et al., 1995) and Recurrence Quantification Analysis (RQA) (Marwan, 2008; Trulla et al., 1996; Zbilut and Webber, 1992). However, Bryce and Sprague (2012) reported negatives of DFA such as the introduction of uncontrolled bias, computational expensiveness and ~~mainly~~ highlight~~ing~~ that DFA cannot provide a generic protection against the nonstationarities of the signals.

2.4.2 Data length

Many of the nonlinear tools are sensitive to time series length (Caballero et al., 2014). For example, given that Multiscale Entropy (MSE) is considered a~~s~~ statistical measure, the data lengths are recommended to be larger to ensure enough samples for the analysis (Costa et al., 2007). Also, LyE (Wolf et al., 1985) and DFA (Peng et al., 1995) metrics are sensitive to data length, while SampEn (Rhea et al., 2011) and FuzzyEn (Chen et al., 2007; Richman and Moorman, 2000) are less sensitive ~~the~~ to time series length. However, the metrics of RQA (Riley et al., 1999; Webber and Zbilut, 1994; Wijnants et al., 2009) and PE (Zunino et al., 2009) are also less sensitive to data length. [how big?]

2.4.3 Sampling rate

One solution when dealing with data length problems is ~~the~~ to increase or decrease of sampling rate (Caballero et al., 2014). However, Duarte and Sternad 2008, p. 267 stated "the increase of sampling rate frequency would only increase artificially the data points without adding information" which therefore ~~is~~ raises the problem of oversamping signals. Then, Rhea et al. (2011) investigated the influence of sampling rate in three entropy measures (ApEn, SampEn and RQAEn) concluding that Ap and RQAEn were robust across to the increase of sampling frequency, while SampEn presented significant difference across all sampling frequencies. Rhea et al. (2011) noted that SampEn is more sensitive to coliniarities than Ap and RQAEn at higher frequencies which lead to a decrease of SampEn. Hence, Rhea et al. (2011) concluded that ~~because~~ signals at higher frequencies appear to be more regular due to the increase of data, therefore producing erroneous entropy results. ~~Then~~ Caballero et al. (2013) showed the robustness of SampEn and DFA tools when using different sampling rate frequencies, stating that frequencies near the dynamics of the activity create a more reliable analysis of the dynamics using DFA values and tested the statement of Duarte and Sternad (2008) that increasing the sampling rate do not increase the gain of information. Caballero et al. (2013) also stated that the decrease of sampling rate frequency is recommended because it presents less consumption of computational power.

2.4.4 Noise

Another point to consider is the noise in ~~of~~ the signals and how such noise affects nonlinear tools metrics (Caballero et al., 2014). For instance, Rosenstein et al. (1993) tested the robustness of LyE against three levels of noise (lowest, moderate and highest) noting the unreliability of LyE exponents in hight-noise environments. However, such cases of unreliability of ~~the~~ LyE is unreal as the reported values of signal-to-noise ratios are substantially ~~s~~lower than the used at the experiments of Rosenstein et al. (1993). Another example~~s~~ is the work of Chen et al. (2009) who compared the robustness of FuzzEn, ApEn and SampEn metrics against different levels of noise, concluding that for a large value of the parameter r of ApEn and SampEn, these two metrics can work ~~fine~~ well with high~~est~~ level of noise, however when noise increases, ApEn and SampEn fail to distinguish time series with different level of noise, whereas FuzzEn ~~probe~~ is ~~to be~~ robust to such highest levels of noise. Also, Bandt and Pompe (2002) by proposing ~~the~~ Permutation Entropy ~~metric~~ (PeEn) ~~showed the~~ to be robust~~ness of PeEn~~ against observational ~~noise~~ and dynamical noise.

Regardless of the source of noise which can be either mechanical (due to recording equipment) or physiological (due to different neural noise), Rhea et al. (2011) highlighted the importance of the effects of noise ~~in~~ on three entropy measures (ApEn, SampEn and RQAEn) which resulted in different results. For instance, values for AnEn and SampEn tended to increase as noise was added to the signals, while RQAEn showed an inverse effect, e.g. RQAEn values decreased as noise in the signal was increased. Similar results for synthetic data were also reported by Pellecchia and Shockley (2015) where RQAEn values decreased from ($RQAEn \approx 5$) for Lorenz system to a ($RQAEn \approx 2$) for a periodic signal with a further decrease ($RQAEn \approx 0.3$) for a sinusoid signal with superimposed noise. Therefore, RQA can also be affected by noise (Rhea et al., 2011). However, the effects ~~with regard to~~ of noise and ~~also~~ nonstarionarity can be mitigated with the selection of the right parameters to perform RQA, particularly, using embedding dimensions from 10 to 20 ~~for biological systems~~ (Webber and Zbilut, 2005).

Another solution ~~in order~~ to deal with noisy time series is the use of traditional filtering methods. However, the attenuation of all frequencies of the signal along the with the noise, given a cutoff frequency, can cut out information that might be useful for nonlinear time-series. Another option is apply DFA, which additionally to the remove of local trends, it also reduces the noise of the signal (Hausdorff et al., 1995). Alternatively, filtering strategies for nonlinear time-series data can be applied which tailor in a more effective way the properties of nonlinear dynamics (see Bradley and Kantz 2015 and references therein).

[The chapter just stops... you should a 'conclusions' paragraph that states the main ideas that are to be used in the thesis] In particular you could why defining embedding parameters is a challenge