Objective and subjective relationships of non-linguistic sounds

---

Physiology Honours 2013

Student

Thomas F Burns

Supervisors

Ramesh Rajan

Yasamin Mokri

Department of Physiology

Monash University

Australia

Contents

I.	Statement of originality	6
II.	Acknowledgements	6
III.	Abstract	7
1.	Introduction	9
1.1	Purpose.....	9
1.2	The need for a new hearing test	9
1.3	Classification of non-linguistic sounds (NLSs)	11
1.4	Objective measures which allow perception of sound attributes	13
1.4.1	Pleasantness	14
1.4.2	Complexity.....	15
1.4.3	Other percepts	17
1.5	Aims and hypotheses	17
2.	Methods.....	19
2.1	NLSs database.....	19
2.2	Psychophysics	19
2.3	Objective measures	19
2.3.1	Selected objective measures.....	20
a)	Sample entropy	20
b)	Permutation entropy	21
c)	Lempel-Ziv algorithmic complexity	22
d)	Duration	23
e)	Peaks-related measures	23
f)	Fractal dimension estimations.....	24

g)	Mean spectral centroid	26
h)	Root mean square of frequency ranges	26
i)	Harmonics-to-noise ratio.....	27
j)	Spectral flatness measure	27
2.3.2	Unselected objective measures	29
2.4	Data analysis	30
2.4.1	Determining salience of objective measures and independentness of the percepts and accuracy of naming.....	30
2.4.2	Linear regressions	30
2.4.3	NLS database homogeneity	31
2.4.4	Multivariate analyses	31
a)	Multiple linear regression and principle component analysis.....	31
b)	Agglomerative hierarchical clustering	32
3.	Results.....	35
3.1	Inter-relationships between percepts and relationship to accuracy of naming NLSs.....	35
3.2	Identification and selection of salient objective measures for NLSs	36
3.3	NLSs database homogeneity	45
3.3.1	The spreads of accuracy of naming, complexity, and pleasantness across different categories of NLSs.....	46
3.3.2	The spread of salient measures across different categories of NLSs	50
3.4	Multivariate analysis	55
3.4.1	Multiple linear regression: paired measures	55
3.4.2	Multivariate relationships	58

a)	Multiple linear regression	58
b)	Principal components analysis	58
3.4.3	Agglomerative hierarchical clustering	61
3.4.4	Agglomerative hierarchical clustering of principal components	66
4.	Discussion	72
4.1	Interactions among percepts and accuracy of naming	72
4.2	Interactions among objective measures	76
4.2.1	Interactions among temporal measures	76
4.2.2	Interactions among spectral measures.....	77
4.3	Sound source differences	77
4.3.1	Accuracy of naming of different categories of sound and differences from deaf people .	77
4.3.2	Differences in the perceptions of different categories of sounds	79
4.3.3	Implications for the training of CI users to perceive NLSs.....	80
4.4	Complexity, familiarity and accuracy of naming.....	80
4.5	Pleasantness	82
4.6	Limitations	83
4.7	Conclusions.....	84
5.	References	86
	Appendix.....	97
	Appendix I : List of sounds and assigned categories.....	97
	Appendix II : Considered temporal and spectral measures of complex waveforms.....	101
	Appendix III : Category make-up of resultant clusters for AHC analyses	104

I. Statement of originality

I declare that this thesis contains no material which has previously been submitted for another degree or diploma in any university and, to the best of my knowledge and belief, this thesis contains no material which has previously been published or written by another person, except when due reference is made in the text of the thesis.

Thomas F Burns

II. Acknowledgements

I wish to thank my many academic mentors and teachers for their assistance throughout my study at Monash University. However I wish to especially thank my academic supervisors for my honours project this year, Associate Professor Ramesh Rajan and Dr Yasamin Mokri, as well as the Physiology Honours Convenors. I am also very grateful for the support offered to me by various technical, research, and administration staff within the Department of Physiology. The friendship and support of post-graduate students and my peers has also been of tremendous value this year, as well as the continual support of my family.

III. Abstract

Our hearing is used to understand speech as well as non-linguistic sounds (NLSs; car horn, laughter, etc.), but how we perceive them and how we correctly identify them appears to differ from the processes used in speech perception. We examined how differences in objective features of different NLSs influence their subjective perception and correct identification.

To do this we utilized a database of 158 NLSs which had been characterised by 12 normal-hearing young adults for the subjective perceptions of complexity, pleasantness, and familiarity, and for accuracy of identification. Using various mathematical, acoustical, and statistical approaches from studies of complex waveforms, we extracted 15 objective features from the NLSs. These objective features were then used to correlate with the subjective perceptions and the accuracy of naming for those NLSs. We additionally used a series of multivariate techniques to establish whether combinations of objective features for NLSs could also account for their differences in subjective perceptions or accuracy of naming. Finally, we also noted the categories of NLSs by their source and used ANOVAs to find if significant differences could be found between these categories for objective features, subjective perceptions, or accuracy of naming.

Significant ($p < 0.05$) inter-relationships between the perceptions and accuracy of naming were identified, which allowed us to determine familiarity as the least independent subjective perception in our study, i.e., it was highly correlated with the other perceptions and accuracy of naming. Significant ($p < 0.05$) inter-relationships between several objective measures were also found, allowing us to identify the salient objective measures. We found several salient objective measures which significantly ($p < 0.05$) correlated (individually) with the perceptions and accuracy of naming for NLSs in our database. Further, combinations of these measures correlated even more significantly; however there were limits to the amount of variation which could be explained with our set of objective measures. Nevertheless, given the objective information of the NLSs, our database could be mathematically categorised into clusters of objectively similar NLSs. These resultant clusters had significant ($p < 0.05$) perceptual differences for complexity and significant ($p < 0.05$) differences in their accuracy of naming. Subjective and objective differences were also noted among different source-

based categories of NLSs in the entire database, though there were fewer significant ($p < 0.05$) differences among these categories than there was potential for.

Our study marks one of the first and most comprehensive efforts to relate objective features to the subjective perceptions of complexity and familiarity in NLSs, as well as adding substantially to our understanding of the same relationships for pleasantness and for identification accuracy. These findings have implications for the development of novel clinical audiometric testing by providing a well-characterised database of NLSs to test a patient's ability to process spectral and temporal components of sound.

1. Introduction

1.1 Purpose

The general purpose of this study lies in the context of developing a novel type of audiometric test to evaluate the consequences of hearing loss. However, as this parent-project is only in its preliminary stages, my project does not probe hearing loss directly. Instead, the aim of my project is to determine what, if any, objective measures of sounds (e.g., the 'peakiness' of a sound's waveform) align with human perceptions of those sounds (e.g., how 'pleasant' a sound is). Specifically, I am interested in non-linguistic sounds (NLSs) or environmental sounds - e.g. car horn, music, laughter - as these lack the confounding semantic contents of language but are still complex - they aren't as simple as pure tones. The human percepts of interest in our study are sound complexity, pleasantness, and familiarity, in addition to sound identification success. These percepts were chosen due to being used in previous research (Marcell et al., 2007). However, before I describe my project in more detail, I wish to provide its appropriate context insofar as describing the rationale behind the parent-project - the need for a new, novel hearing test.

1.2 The need for a new hearing test

Hearing loss is estimated to affect 250 million people worldwide and is the most common sensory deficit in humans (Mathers et al., 2000). Given that the inner ear effectively segregates all sounds into their basic frequency content and that this segregation is maintained in the tonotopically-organised flow of information from cochlea through at least seven way-stations to cortex, the simplest and easiest way of determining hearing loss is audiometric measurement of hearing thresholds at individual sound frequencies (WHO, 1991). However, while such an audiometric test is useful in identifying hearing loss and the pattern of inner ear damage, it is limited (except through experience of the audiologist) in predicting the functional consequences of the hearing loss, i.e., how that specific pattern of hearing loss impacts on a person's ability to understand speech or complex NLSs that make up a normal soundscape. Thus, other types of audiometric tests using complex sounds can reveal functional hearing losses which are otherwise left undetected (Possamai et al., 2012).

The most common of these complex sounds tests are speech tests, especially speech-in-noise tests (Jayaram et al., 1992; Besser et al., 2012; Wilson et al., 2012). As these tests are simple to conduct and provide clinicians with basic as well as ecologically valid information about a patient's hearing ability, they are now considered the apex of the clinical armamentarium (Wilson, 2011). However, these tests ignore a substantial portion (Ballas, 1993) of the everyday human soundscape: environmental sounds or NLSs, which are not well perceived in some pathologies (Looi & Arnephy, 2010). NLSs hold importance both pragmatically (e.g., horn of an approaching car, the sound of a fire alarm) and emotively (e.g., Beethoven's ninth symphony, the cry of a baby), and the ability to recognise NLSs when there is a hearing loss is not directly predicted by other functional tests like speech tests or even speech-in-noise tests. Thus, Inverso & Limb (2010) showed that accurate language perception does not equate to accurate NLS perception. In their study, cochlear implant (CI) patients underwent various speech and NLS identification tests. Of the participants who scored >80% words correct on a standard speech test (n=16), their ability to correctly identify the category of NLS was similar to their speech perception scores (76.1% correct) whereas their ability to identify the actual NLSs themselves was much less accurate (53.9% correct). This may be due to NLSs or features thereof activating distinctly different parts of the brain than speech (Lewis et al., 2004; Meyer et al. 2005; Leaver & Rauschecker, 2010). Further, different categories of NLSs seem to activate unique combinations of brain areas (Lewis et al., 2005; Doehrmann et al., 2008; Leaver & Rauschecker, 2010; Leech & Saygin, 2010). There is also the fact that NLSs carry meaning and affective value as well as content (a jackhammer being considerably less loaded with emotional value than the sound of a child crying) and this information may be critical in identifying a NLS; it is unknown if the retrieval of such information is affected by hearing loss as they are in some cognitive disorders (Crandell, 1991; Akeroud, 2008; Rudner et al., 2009; Choi et al. 2011).

Thus, for three basic reasons it is important to develop a test of complex non-linguistic everyday sounds: (1) testing a person's ability to accurately perceive NLSs will not only be testing fundamentally different stimuli but will also be testing unique brain areas. It is worth noting that adding to its potential value as a clinical tool is the fact that such a test would lack the potentially confounding aspects of language present in any speech-based tests, making it useful for children and

listeners with poor language skills or for whom a hearing test in their native language is unavailable; (2) NLSs constitute a set of very important environmental sounds that have importance for survival let alone the richness of life and NLSs carry value beyond their content in having an affective value (the soothing sound of flute music or the annoyance from chalk being rubbed down a blackboard); and (3) recognition of NLSs is not predicted by other current tests of hearing loss.

1.3 Classification of non-linguistic sounds (NLSs)

A NLS-based hearing test needs to be ecologically representative of the soundscape and therefore useful in predicting the functional consequences of hearing loss across a variety of the daily scenarios that is represented by this broad set of sounds. Hence, the test must contain sounds from different categories that are unique from one another. This raises the question of how this broad set of NLSs can be classed, or if they can be grouped together at all.

One approach, and the simplest, is to categorise NLSs based on the source of the sound or the imagery evoked by the sound. Thus, Leaver & Rauschecker (2010) categorised 26 NLSs into “songbirds”, “other animals”, and “musical instruments” (clearly a very restricted range of sounds and categories) and Inverso & Limb (2010) categorized 50 NLSs into “animal”, “human non-speech”, “mechanical/alerting”, “nature” and “musical instruments”. The distinctions between the latter classes are clearly hazy – “animal” and “nature” categories could both subsume a bird song; “human non-speech” and “alerting” could be equally applied to classify a warning yell. This ambiguity is well illustrated in a study by Marcell et al. (2007) in which 38 participants were allowed to freely classify 60 NLSs and the classes were then grouped as much as possible for similarity by two independent judges. This yielded 27 categories of sounds (four-legged animal, accident, air transportation, bathroom, bird, farm animal, game/recreation, ground transportation, human, hygiene, insect, kitchen, musical instrument, nature, pet, reptile/amphibian, sickness, signal, sleep, tool, water/liquid, weapon, weather, household, paper, machine, and other) - clearly an unworkably large classification system for any test battery but illustrative of the problems with trying to classify NLSs based on their source or imagery. Further it is clear that even this large imagery-based classification is far from flawless. In

the same study (Marcell et al., 2007), another 49 participants were asked to classify the same 60 sounds but constrained to the 27 categories generated from the first group. It was found that there could be disagreement within the new batch of participants – for example, 61% of these participants classed a duck sound into the “bird” category while 36% of them classed the same duck sound into the “farm animal” category (Marcell et al., 2007). It may be such reasons that led to Gygi et al. (2007) formulating a classification system which combined this basic image/source classification with classification based on acoustic properties of sounds – they classified 50 sounds into the main categories of (a) harmonic sounds (sub-categories: alerting, human, animals, or ringing), (b) discrete impact sounds (sub-categories: single impact or multiple impact), and (c) continuous sounds (sub-categories: water, scraping, unvoiced vocalisations, or mechanical).

A problem with all of these classification systems is that they constrain the use of the NLSs purely to tests of identification and do not allow evaluation of any other percept, especially those that span different image-based categories of NLSs (e.g., different sounds may all be perceived as equally pleasant or equally complex). As noted by Marcell et al. (2007 p. 561) *“Previous research has focused on the identification of relatively brief, isolated environmental sounds; in contrast, less is known about how people perceive and understand complex, coherent, and lengthy sequences of natural sounds, such as the sounds accompanying the feeding of a cat or the fixing of a leaky faucet”*. Gygi (2001) noted that identification of a NLS may require the use of different objective measures of the sound than if some other attribute was requested : *“if the goal is to identify the source of a particular sound, the listener may have to focus on short-term spectral-temporal properties (to identify it as rapidly as possible) as opposed to tracking over an extended period of time, which is required to extract the ‘message’ conveyed by speech or music”* (Gygi, 2001, p.5).

In an attempt to go beyond simple identification of NLSs, Marcell et al. (2007) asked their subjects to categorise NLSs by their perceived complexity, pleasantness, and familiarity. This allows testing of whether other aspects of perception of an NLS are affected by hearing loss or cognitive impairment – e.g., whether the affective value of an NLS is changed. Unfortunately a significant limitation of this classification is that if a test is to use these NLSs, it is constrained to the sounds for which the classification on the basis of familiarity, pleasantness, and complexity has already been

obtained. This is a particular problem when it is noted that the Marcell et al. (2007) study used a large number of participants (n=124) but tested only 22 “sound events” and it is unlikely that such a limited set can be truly representative of a normal soundscape or that such a limited set is even enough to conduct a full test battery. The problem is exacerbated by the fact that each time it is desired to add a new sound to the test battery, the entire process of testing this new sound for these percepts of familiarity, pleasantness, and complexity has to be repeated with a large number of subjects.

1.4 Objective measures which allow perception of sound attributes

An alternative approach for the classification of NLSs is to determine if the perception of NLS attributes aligns with particular *objective features* of sounds, i.e., regardless of the source of two NLSs, do similar spectral or temporal features of the NLSs determine how they are perceived for complexity, or familiarity, or pleasantness, or how successfully they are identified? We know from our own experience that we can perceive a flute and a birdsong as pleasant and yet their sources are clearly very different. Finding objective measures of NLSs that align with a particular percept of those sounds has another great advantage: a test of NLSs for that percept can select NLSs to span a range of those objective measures without having to first test each NLS for that percept. New sounds could be added to the test battery, or sounds grouped together for a battery requiring sounds of a similar value of the test percept (e.g., similar level of complexity) based on these objective measures. The latter could then allow for tests of another percept with sounds “normalized” for a potentially confounding percept – e.g., tests of pleasantness may be able to use NLSs with similar complexities so that the latter property was not a confound. Finally, if it can be found that particular objective measures of sounds align with specific percepts it would provide remarkable (and hitherto unavailable) information about what types of processing are affected in hearing loss or cognitive impairment and this may have powerful value for diagnostics or therapeutics – for example, the cochlear implant can, in very many cases, allow full or near-full restoration of the ability to understand even open-set (effectively, previously unheard) sentences yet does not restore the ability to understand or appreciate music (Nimmons et al., 2008). Clearly, in this case, the processing restored by the implant even for

something as complex as speech does not restore the ability to process the objective measures that allow processing of music, though at face value a significant number of the properties are the same or similar in the two types of sounds, speech and music (Wolfe, 2002).

Classifying NLSs into meaningful perceptual classes requires the listener to correctly distinguish class-unique feature-sets which include a very diverse array of objective measures (Ballas, 1993). Thus, it is not a trivial task to determine what objective measures actually do align with the percept. For example, even for simple identification, for most sounds fine spectral resolution is critical (Shafiro, 2008). However, some sounds actually become more difficult to correctly identify with increased spectral resolution (Shafiro, 2008) as they rely more heavily on temporal properties. For such sounds, classification is not instantaneous and requires a reasonable time exposure in order to gain the required information, though this varies between sounds (Ballas et al., 1986; De Lucia et al., 2012). It is therefore not surprising that Marcell et al. (2007) reported that for 120 sounds, there was no correlation between duration of sound and either the accuracy of subjects' naming responses or their level of confidence in the naming responses, suggesting that a combination of temporal and spectral information is needed for humans to correctly classify sounds according to their features. This may explain the fact that there has only been limited study of objective measures of NLSs. I now review these studies and establish that while some of these studies attempted to align objective measures of sounds with a human percept, many do not use a sufficiently large battery of sounds to allow for generalised conclusions and much remains to be known about particular percepts, including their interactions (e.g., does complexity determine pleasantness).

1.4.1 Pleasantness

Although the percept of pleasantness is often associated with nature or music sounds (Shimai et al. 1993), the underlying properties responsible for this are less well reported, though they appear to be spectral-based (Halpern et al., 1986; Kumar et al., 2008; Reuter & Oehler, 2011). The first study to conduct significant work on this topic was by Halpern et al. (1986), who took recordings of 16 NLSs and then re-synthesized the sounds with 11 different high- or low-pass filters. Twelve listeners then rated how pleasant or unpleasant they perceived the sounds to be. Contrary to the experimenters'

intuition, it was found the low- and mid-range frequencies, not the high-range frequencies, were responsible for eliciting unpleasantness in the sounds. This supported later work by Kumar et al. (2008), who collected unpleasantness ratings from 50 listeners for 75 sounds and showed that high energies in the frequency ranges of 2500-5500Hz and temporal modulations within 1-16Hz were responsible for perceptions of greater unpleasantness. Further work by Reuter & Oehler (2011), who replicated the Halper et al. (1986) study but conducted a more slightly advanced analysis, concluded that high energies at frequencies between 2000-4000Hz caused unpleasantness.

As an interesting aside, similarities between the properties of unpleasant sounds and macaque monkey warning cries has led to the hypothesis (Halper et al., 1986) that the original functional purpose of unpleasantness as an auditory feature is as a means of communicating distress or warnings. However this similarity has not been quantitatively tested and other emotive, affective, or audio-haptic interactions may be involved (Zald & Pardo, 2002; Cox, 2008; Kumar et al., 2012).

1.4.2 Complexity

The complexity of NLSs is occasionally implied or alluded to (Lewis et al., 2004; Gygi et al., 2007) but rarely explicitly quantified (Jayant & Noll, 1984; Boersma, 2001; Shibayama, 2006; Xue et al., 2008; Singh, 2011). This gap in the literature is compounded by the ambiguity of what different fields of study describe as ‘complex’, e.g. applied mathematics and computer science tend to favour information theory - how much unique information is contained in something (Hartley, 1928; Shannon, 1948; Kolmogorov, 1968) - whereas pure mathematics and abstract applications favour chaos theory - how perceivably unpredictable sequential information appears (Lorenz, 1963; Mandelbrot, 1963).

In psychoacoustics, complexity has generally been quantified or inferred by acoustical and mathematical means. One of the most basic acoustic measures identified as potentially describing complexity of a sound was the spectral flatness measure (SFM), also known as the Wiener entropy (Jayant & Noll, 1984; Boersma, 2001; Dubnov, 2004). Used in birdsong research (Tchernichovski et al., 2000) and in audio compression applications (Johnston, 1988), the SFM is a simple ratio of the geometric mean by the arithmetic mean of a sound’s power spectrum. Essentially, this identifies how

tonal or noisy the sound is, with the classification being that the more tonal a sound is, the 'simpler' it is. However this creates the disadvantage of ignoring temporal complexity and to address this, a variation of the SFM called the spectral structure index (SSI) or variation (SSV) (Reddy et al., 2009; Singh, 2011; Lewis et al., 2012; Talkington et al., 2012) incorporates temporal information by comparing the SFM in sections of a sound against each other to see how the SFM varies over time. This has been suggested as a direct means of implying complexity of a sound (Singh, 2011), as well as used to correlate with cortical processing and subjective distinctness in NLSs (Lewis et al., 2012; Talkington et al., 2012). However, when used in isolation, this ignores the broader context of the spectral and even overall macro-temporal information.

Another measure which has been proposed to directly imply complexity is Fractal dimension estimation (Shibayama, 2006). From the field of geometry (Mandelbrot, 1967), this measure indicates the level of intricate self-similarity of a shape, e.g. a waveform. In this way, it focuses on the temporal features of a sound. However it does not necessarily ignore its spectral contents inasmuch as it inherently detects temporal patterns or similarities. In the case of a one-dimensional signal, this could theoretically identify micro-periodicity in a sound's waveform.

It must be noted that complexity, as defined in these studies, is acoustical or mathematical complexity; none of these studies have attempted to relate these objective complexity measures back to subjective perception of NLS complexity. The same is true for our final percept of interest, familiarity; we could find no studies which linked the percept of familiarity to any objective features of a sound. Despite this knowledge gap, familiarity of NLSs has been shown to affect subsequent auditory processing (Cycowicz & Friedman, 1998; Kirmse et al., 2012).

Some studies (Ballas, 1993; Kidd & Watson, 1999; Gygi et al., 2004) have attempted to relate objective properties of NLSs back to the difficulty of their identification. However these objective properties were not very broad (Ballas, 1993; Kidd & Watson, 1999; Gygi et al., 2004), mostly determined qualitatively via spectrograms by a single observer (Ballas, 1993), or were based upon artificially modifying the sounds instead of directly measuring a specific quality (Kidd & Watson, 1999; Gygi et al., 2004).

1.4.3 Other percepts

Many other types of percepts have been identified in NLSs and some have been related to objective measures. Although the main percepts of interest in our study do not include any of these, a brief survey of these percepts is included here in the interest of completeness and to highlight the importance of selecting both spectral and temporal objective measures in any attempt to map objective measures to subjective percepts. Despite this importance, the objective measures which have been used to relate to various other sound percepts have often related to spectral components, with limited work considering temporal components.

As examples: unique combinations of sound frequencies have been shown to allow a listener to identify a material being struck as wood or metal (Pollack & Ficks, 1954; Gaver, 1986, Graham, 1999), and energy at high frequencies and spectral peak information allow identification of a walker as either male or female (Li et al., 1991). Other spectral characteristics, such as a variation of spectral energy distribution and fluctuation, as well as harmonic structure, have been implicated in the perception of musical timbres (Grey, 1977; Grey & Grodon, 1978). Some percepts, such as the implied level of urgency in an alarm, have been shown to relate to spectral components like inharmonicity and pitch range and to temporal components like speed (relative to a perceived natural speed) and melodic structure (Momtahan, 1990; Edworthy et al. 1991; Hellier et al., 1993; Burt et al., 1995; Haas & Casali, 1995; Haas & Edworthy, 1996). Conversely, other studies have found that some percepts rely primarily on temporal components, including the effect of impulse synchronicity and amplitude on the perception of whether a glass bottle cracks, breaks or bounces (Warren & Verbrugge, 1984).

1.5 Aims and hypotheses

It is obvious from this introduction that there has been almost no satisfactory or complete attempt to determine if objective measures of NLSs align with the way humans perceive those sounds for various fundamental percepts such as complexity, pleasantness, and familiarity. My study fits into

this gap, and I have aimed to study whether objective measures of NLSs align with percepts of interest and identification accuracy of publically available NLSs.

Therefore, our formal hypotheses are that:

1. The subjective percepts of complexity, pleasantness, and familiarity in NLSs within our database, and their ability to be accurately identified, can be quantitatively described by and correlate with objective measurements of their temporal or spectral contents; and
2. Different classes of NLSs within our database possess unique feature-sets of objective measures which are related to their perceptual differences or their rates of identification success.

2. Methods

2.1 NLSs database

Publically available NLSs used in previous research (Marcell et al., 2007) as well as NLSs from online multimedia archive sources (Shafiro & Gygi, 2004) were combined to create a database of 158 sounds. Within this database, 144 sources with 14 exemplars existed across a broad range of categories (Appendix I). All sounds were normalised for amplitude.

2.2 Psychophysics

Previous work in the laboratory conducted by Rahman (2011) involved recruiting 12 normal-hearing observers (mean age 19.8 years, SD 0.57) from the student population at Monash University, Australia, and testing their hearing using audiometry to ensure normal thresholds for their age. In groups of three, these participants listened to the NLSs individually via speakers in a quiet room. After listening to each sound once, the participants were instructed to rate the sound on seven-point Likert scales for ‘complexity’, ‘pleasantness’ and ‘familiarity’ (percepts), and also asked to name the sound source (accuracy of naming). This mimicked the method which was used by Marcell et al. (2007). Ethics approval for the collection of this data was obtained from the Monash University Standing Committee on Ethics in Research in Humans. The raw data from this work (Rahman, 2011), was provided to me digitally and formed the entirety of my psychophysical dataset for this project.

2.3 Objective measures

A wide range of temporal and spectral measures of complex waveforms from various disciplines were considered for inclusion in this study (Appendix II). Particular emphasis was placed on those which measured complexity or have been shown to be related to sound percepts relevant or identical to those used in this study. The selected measures were then calculated for the entire NLSs database using a commercial software package (MATLAB R2012a, The MathWorks Inc., Natick, MA, United States) and a phonetics research program (Boersma, 2001).

2.3.1 Selected objective measures

The measures which were selected for our study were chosen for differing reasons; these reasons and their method of calculation are described for each selected measure below:

a) Sample entropy

Entropic measures were considered for our study due to their established use in implying the complexity of biological signals such as electroencephalography (Rezek & Roberts, 1998; Bruhn et al., 2000), respiratory motion (Burioko et al., 2003), and heart rate (Lake et al., 2002). However, such measures can rely on different views of theoretical 'complexity'. The two most prominent theories which entropic measures are based upon is information theory (how much unique information is contained in a signal; Hartley, 1928; Shannon, 1948; Kolmogorov, 1968) and chaos theory (how perceivably unpredictable sequential information appears; Lorenz, 1963; Mandelbrot, 1963). It is therefore reasonable to include two entropic measures, one based on each theoretical perspective of complexity.

From those based on information theory, we chose sample entropy (Richman & Moorman, 2000; Lake et al., 2002) as it is based on the well-known approximate entropy measure (Pincus, 1991; Rezek & Roberts, 1998; Bhattacharya, 2000; Burioka et al. 2003) but lacks its self-matching bias (Richman & Moorman, 2000). The measure estimates entropy as the likelihood of two identical strings of some length being similar for the next number in the sequence, therefore identifying the amount of repeatable or non-unique information. An added option to our choice of sample entropy was the use of multi-scale and multi-resolution techniques (Costa et al. 2003; Costa et al. 2005; Torres & Camero, 2000), which would help avoid oversampling or signal artefacts. However, because our signals (NLSs) had been vetted for artefacts, i.e., they were high-quality recordings, and are relatively brief in duration, the option to conduct our sample entropy analyses including these techniques was considered unnecessary.

Therefore, to calculate the sample entropy (Richman & Moorman, 2000) of a NLS, we consider its waveform which has a constant sampling time to have a length of N in a single embedding dimension of $m = 2$, so as to give a dataset of $N = \{x_1, x_2, x_3, \dots, x_N\}$. A template vector was created

with a length of m to map $X_m(i) = \{x_i, x_{i+1}, x_{i+2}, \dots, x_{i+m-1}\}$, as well as a distance function to determine distance within the vector space, e.g. Chebyshev distance $d[x_m(i), x_m(j)]$ where $(i \neq j)$. Next, we determined a standard tolerance r of $0.2 \times std(m)$, where $std(m)$ was the standard deviation of m from $1 \rightarrow N$. Given r , we then counted the template vector pairs of length m (which we call B) and $m + 1$ (which we call A) wherever $d[x_m(i), x_m(j)] < r$. We then calculated the sample entropy as $SampEn = -\log \frac{A}{B}$, where the higher a value, the more unique information present throughout the NLS and vice-versa.

b) Permutation entropy

Having already selected an entropic measure based on information theory, we then chose another entropic measure based on chaos theory. Of those which exist in the literature, we chose permutation entropy (Bandt & Pompe, 2002) as it is one of the best recognised (Riedl et al., 2012; Zanin et al., 2013). The measure works by calculating the distribution entropy of permutations of small bins of sequential numbers within the dataset. We again had the option of considering multi-scale and multi-resolution techniques (Costa et al. 2003; Costa et al. 2005; Torres & Camero, 2000), however as our signals were still high-quality recordings and relatively brief in duration, we again considered these techniques unnecessary for our dataset.

To calculate the permutation entropy (Bandt & Pompe, 2002) of a NLS which has a constant sampling time and a length of n , we constructed an embedding dimension of $D = 4$ at each time-point s of the D -th subsequent values in the waveform $X = (x_t: 1 = 1, \dots, n)$ but with a given time delay of $\tau = 50$ (in data-points), i.e. $s \mapsto (x_{s+\tau+1}, x_{s+\tau+2}, x_{s+\tau+3}, x_{s+\tau+4})$. An ordinal pattern was then assigned to this created vector as the permutation $\pi = (r_0 r_1 r_2 r_3)$ of $(01 \dots D - 1)$, ensuring that $x_{s+r_0} \leq x_{s+r_1} \leq x_{s+r_2} \leq x_{s+r_3}$. We then determined the permutation entropy by applying the Shannon entropy (Shannon, 1948) given a probability distribution of Π for the entire series and considering the frequencies of possible permutation patterns i as elements thereof for $i = 1, \dots, D!$ and defining $PermEn = -\sum_{i=1}^{D!} \pi_i \log_2 \pi_i$.

c) Lempel-Ziv algorithmic complexity

Yet another way of thinking about complexity is algorithmic or computational complexity theory - how much the original information can be condensed into a computationally-efficient representation, as in an algorithm. Measures based on this theory, such as state machines, have been used in birdsong research with some success (Sasahara & Ikegami, 2004), however this particular measure requires that signals be highly repetitive and so our dataset was not suitable. Another measure based on this theory which has been described as relating to complexity is the measurement of hidden Markov model states (Gunther, 1994), however this has not always been considered an explicit measurement of complexity and is more often used for classification purposes (Shalizi et al. 2001; Shalizi et al. 2002; Buchler et al. 2005). One of the most popular and successful measures based on this theory, though, has been the Lempel-Ziv (LZ) algorithmic complexity measure (Khalatur et al. 2003; Szczepanski et al. 2003; Watanabe et al. 2003; Huang et al. 2003), though it was originally designed as a lossless digital compression algorithm (Ziv J & Lempel A, 1978). It works by first reducing the complex signal into a list of symbols, and in physiological data it has been shown that binary sequences are sufficient for this purpose (Xu et al., 1997; Radhakrishnan & Gangadhar, 1998; Zhang et al., 2000; Wu & Zu, 2001; Zhang et al., 2001). Then, moving from left to right along the signal, we count how many unique instances of consecutive symbols (or, in our base, ones and zeros) exist in the entire signal. We then normalise the signal according to its length. In the absence of hard-and-fast rules regarding the method of reducing a complex signal to a binary sequence, we opted to use some of the most common methods, including: (i) average binary: a one is assigned to any data-point above the average within the dataset, and the remainder are assigned a zero; (ii) modified zone binary: any data-point outside of the 'modified zone' (the average within the dataset $\pm$ the standard deviation) is assigned a one, and the remainder are assigned zero; and (iii) differential binary: where a data-point is positive with respect to that which was immediately before it, a one is assigned (including the first data-point), and where it is negative a zero is assigned.

Having reduced the signal to a binary form using the three abovementioned methods, we then calculated the LZ algorithmic complexity (Aboy et al., 2006) for each of these three sequences separately for each NLS' waveform. For each sequence P , subsequences S and Q were created and the

last character deleted π from Q , making $SQ\pi$ the new total sequence. We assigned $v(SQ\pi)$ as the 'vocabulary' of unique subsequences of $SQ\pi$ and began counting the unique subsequences at $c(n) = 1$ where $S = s(1)$ and $Q = s(2)$. This continued as $S = s(1), s(2), \dots, s(r)$ and $Q = s(r + 1)$ to give $SQ\pi = s(1), s(2), \dots, s(r)$, except where Q was already found in the vocabulary $v(SQ\pi)$. Where it was found, Q was increased to be $Q = s(r + 1), s(r + 2), \dots, s(r + i)$, i.e. until it was not longer a subsequence of $SQ\pi = s(1), s(2), \dots, s(r + i - 1)$. Each time a new subsequence was found, $c(n)$ was increased by one. After exhausting a subsequence, the subsequences were refreshed as $S = s(1), s(2), \dots, s(r + i)$ and $Q = s(r + i + 1)$ until Q was the final digit in the sequence (the penultimate of the original sequence P). To normalise this complexity measure $c(n)$ of P , we normalised $c(n)$ according the length n of P given the number of used symbols $\alpha = 2$ and determining the upper-bound of $c(n)$ as $b(n) = \frac{n}{\log_2(n)}$, defining the normalised LZ complexity $C(n)$ as $C(n) = \frac{c(n)}{b(n)}$.

d) Duration

It has been well established (Ballas, 1993; Gygi et al. 2007) that sound duration is an important feature of a NLS event, and can have an impact on its identification success (Ballas, 1993; Ballas et al., 1986; De Lucia et al., 2012) and in objective classification (Gygi et al. 2007; Maher & Studniarz, 2012). Given this, and that we were partly interested in identification success of NLSs, we considered duration to be an important measure for inclusion in our study. However, it did not require a formal calculation for our database as the NLSs' durations were already associated for each individual sound.

e) Peaks-related measures

Unlike duration, the use of peaks-related measures has been less widespread in NLS research (Gygi et al. 2007). However, having been included in a prominent and comprehensive previous NLS study (Gygi et al. 2007), inclusion of this measure would allow our study's results to more easily supplement or be compared with this previous study. As the name implies, there are multiple peak-related measures, however all rely on the premise of identifying peaks within a sound's waveform. Because a 'peak' could be highly relative to a particular sound, i.e. there may be moments of relative

silence and relative loudness, we chose to first identify the range of data-points within a sound's waveform, and call any data-point a 'peak' if its amplitude was $> 0.8 \times \text{range}$ where the *range* was the overall amplitude range within the NLS. Given the number of peaks and their relative amplitudes (where 1 is the greatest amplitude) for a NLS, three measures were derived: (i) the number of peaks per second; (ii) the mean relative amplitude of all the peaks; and (iii) the standard deviation of the relative amplitude of all the peaks.

f) Fractal dimension estimations

As we were interested in the complexity of NLS', any measures we could find which inferred or estimated the complexity of a biological signal were considered for inclusion in our study. Of these, entropic measures were the most common and have been discussed earlier. However, another very well-known measure is fractal dimension (FD) estimation (Spasic et al., 2005; Shibayama, 2006; Raghavendra & Dutt, 2010).

Formally, all geometric objects have both a topological and fractal dimension; however the former is always an integer whereas the latter may be an integer or a non-integer. We can define the topological dimension of an object as the number of independent directions it is possible to move within that object, e.g. on a curve we can only move in one direction or the reverse $\therefore$ topological dimension of a curve = 1. Whereas the definition of a fractal dimension is much more complex (Mandelbrot, 1983; Barnsley, 1993) but essentially is a measure of self-similarity within a geometric object after magnification. A well-known fractal (object which has a fractal dimension $>$ topological dimension) is the Koch snowflake (Koch, 1904), which is a curve of infinite length and a topological dimension of 1, however its fractal dimension is $\frac{\ln 4}{\ln 3} = 1.26186\dots$, i.e. it is self-similar to the extent of being complex beyond its topological dimension.

Because NLSs are real signals - i.e. unlike the Koch snowflake, they have a finite length - we are assuming fractal properties are present within them and then estimating what their fractal dimension could be if they were of an infinite length. Of the methods available for fractal dimension estimation of one-dimensional signals, some have analysed artificial signals (Raghavendra and Dutt, 2010) in order to compare their estimations with parametric fractal signals which have known

theoretical fractal dimensions. This has enabled such methods to be tested for their accuracy. Of these methods, the Higuchi method (Higuchi, 1998), which was originally designed to analyse magnetic field data, has become one of the most commonly used for one-dimensional signals. However, a method called the Normalised Length Density (NLD) has also been developed for very short samples or epochs, especially useful where there are 100 data-points or less (Kalauzi et al. 2009). Although both of these methods of FD estimation - Higuchi and NLD - attempt to estimate the same feature of a signal, they do so in very different ways. Because of this difference, we chose to include both methods in our study.

To calculate the FD estimate using the Higuchi method (Higuchi, 1998), we let X be the one-dimensional time series of the waveform for a given NLS $X = x[1], x[2], \dots, x[N]$, where N is the final sample in the series. We then defined a new time series as

$$X_k^m = \{x[m], x[m+k], x[m+2k], \dots, \\ x\left[m + \text{int}\left(\frac{N-m}{k}\right) \times k\right]\}$$

where k was the sampling frequency and m was the initial time point in the series. We then found the length of the new time series X_k^m by

$$L(m, k) = \frac{\left\{ \left(\sum_{i=1}^{\text{int}\left(\frac{N-m}{k}\right)} |x[m+ik] - x[m+(i-1) \times k]| \right) \left[\frac{N-1}{\text{int}\left(\frac{N-m}{k}\right)} \times k \right] \right\}}{k}$$

where N is the length of the original time series X . The length for the time interval k was then defined as

$$L(k) = \frac{1}{k} \times \sum_{m=1}^k L(m, k).$$

To calculate the actual estimate, we used a double-logarithmic scale to plot $L(k)$ against $\frac{1}{k}$, with $k = 1, 2, \dots, k_{max}$, where the slope $\left\{ \ln[L(k)], \ln\left(\frac{1}{k}\right) \right\}$ was equal to the fractal dimension estimate of the original one-dimensional time series X of the NLS. For this study a value of $k_{max} = 10$ was

chosen as after testing $k_{max} = 2, 3, \dots, 50$ it gave us the greatest dynamic range (Spasic et al. 2005) within our dataset of NLSs.

The second method, NLD (Kalaui et al. 2009), is much simpler than the Higuchi method. We let y be the amplitude normalized one-dimensional time series of the waveform for a given NLS $y = y(1), y(2), \dots, y(N)$, where N was the final sample in the series. We then defined the NLD as

$$NLD = \frac{1}{N} \sum_{i=2}^N |y_n(i) - y_n(i-1)|$$

where $y_n(i)$ was the i th sample in the time series. We then used the power model (Kalaui et al. 2009) to convert the NLD to an FD estimate

$$FD = a(NLD - NLD_0)^k$$

where $a = 1.9079$, $NLD_0 = 0.097178$ and $k = 0.18383$.

g) Mean spectral centroid

The mean spectral centroid, similarly to duration, is a well-reported, important feature of NLSs in terms of objective classification (Shao et al. 2003; Gygi et al. 2007; Maher & Studniarz, 2012). It has also been aligned with subjective dimensions of musical instruments (Grey & Grodon, 1978). For these reasons, we chose to include it in our study.

Calculation of the mean spectral centroid was based on the fast Fourier transform (FFT) of the NLS and given as:

$$\omega_c = \frac{\int_0^{\omega_0} \omega |F(\omega)|^2 d\omega}{\int_0^{\omega_0} |F(\omega)|^2 d\omega}$$

where ω_0 is half the sampling frequency of the NLS.

h) Root mean square of frequency ranges

The root mean square (RMS) of frequency ranges was employed by Gygi et al. (2007) in their objective classification of NLSs, and so inclusion of this measure in our study would help provide a comparison. High power at certain low- and mid-range frequencies have also been heavily implicated with the perception of unpleasantness (Halpern et al., 1986; Kumar et al., 2008; Reuter & Oehler, 2011), however these studies have used varying techniques to reach their conclusions.

As conducted in the Gygi et al. (2007) study, we filtered NLSs to isolate certain frequency ranges and then calculated the RMS of those filtered signals. It is important at this point to remember that all NLSs in our database had been normalised for amplitude (RMS) already, so comparing the RMSs of different NLSs was appropriate. However, unlike the Gygi et al. (2007) study, our study rests in the context of a larger study concerned with hearing tests. Because of this, we chose frequency ranges based upon the frequencies commonly tested using audiometry (WHO, 1991). We therefore chose the following frequency ranges to isolate via filtering: 250-2500Hz, 500-1000Hz, 1000-2000Hz, and 2000-4000Hz.

i) **Harmonics-to-noise ratio**

Due to previous studies employing the use of harmonic measures as an important feature of NLSs (Gygi et al., 2007; Lewis et al., 2009), we chose to include one of the most recognised of these measures, the harmonics-to-noise ratio (HNR). Although this measure was originally developed as a psychophysical rating of the hoarseness of human speech (Yumoto et al. 1982), later studies have used it the study of NLS features, e.g. correlating it and other features with functional magnetic resonance imaging of brain areas (Lewis et al. 2009). The HNR works by quantifying the relative power of harmonic contents of a sound as the energy of its periodic contents in comparison to its non-periodic contents (Boersma, 1993). The value of zero represents equal contents, a negative value represents a greater non-periodic contents than periodic contents (e.g. white noise), and vice-versa (e.g. a wolf howl).

To calculate the HNR, we let $f(t)$ be the original one-dimensional waveform for a given NLS and $f_i(\tau)$ be the waves for each pitch period τ . We then defined the HNR as:

$$\frac{H}{N} = \frac{n \int_0^T f_A^2(\tau) d\tau}{\sum_{i=1}^n \int_0^{T_i} [f_i(\tau) - f_A(\tau)]^2 d\tau}$$

where n is the number of pitch periods and d is the duration of the pitch period.

j) **Spectral flatness measure**

Although mostly used in phonetics research (Boersma, 2001), the spectral flatness measure (SFM) was chosen for inclusion in our study due to its unique status as an entropy-related measure but

in the spectral domain and its use in previous sound research (Jayant and Noll, 1984; Boersma, 2001) c.f. most entropy-based measures measuring something in the time domain and not being in widespread use for sound research, e.g. permutation entropy (Riedl et al., 2012; Zanin et al., 2013). The SFM works by measuring the degree of uniformity within a signal's power spectrum (Jayant and Noll, 1984), with white noise = 1 (chaos) and a pure tone = 0 (complete order). However to expand its dynamic range, it is measured on a logarithmic scale so that white noise = $\log 1 = 0$ and a pure tone = $\log 0 = -\infty$. This results in large negative numbers for simple sounds and numbers approaching zero for complex sounds.

To calculate the SFM, we let $s(t)$ be the FFT of the time-series of a NLS and defined the SFM as:

$$SFM = \log \frac{[\prod_{f=1}^N S(f)]^{\frac{1}{N}}}{\left(\frac{1}{N}\right) \sum_{f=1}^N S(f)}$$

where N was the number of FFT points and $S(f)$ was the amplitude (Hz) at each point. The result was then logged (1) to expand the dynamic range.

k) Spectral structure variability or index

Using the derived SFM from a signal, we can then calculate the spectral structure variability (SSV) or index (SSI) as the variability of the SFM over time. It is conceptually similar to how a spectrogram shows the variability of Fourier transforms over time. This measure has been used previously in the analysis of NLSs (Reddy et al. 2009; Singh, 2011) and human vocalisations (Talkington et al. 2012). It has even been suggested as a direct measure of NLS complexity (Singh, 2011). Since we are interested in complexity and this measure has been used to characterise NLSs, we considered it suitable for inclusion in our study.

The SSI gives a pure number, where if $SSI < 1$ the sound has similar spectral dynamics (simple) and if $SSI > 1$ the sound has dissimilar spectral dynamics (complex). To calculate it, we let $s(t)$ be the FFT of the time-series of a NLS and defined the SSI as:

$$SSI = \sum \frac{\left[SFM(t) - \frac{\sum SFM(t)}{N} \right]^2}{N}$$

where N is the number of time frames of:

$$SFM(t) = \log \frac{[\prod_{i=1}^N S(t, f)]^{\frac{1}{N}}}{(\frac{1}{N}) \sum_{i=1}^N S(t, f)}$$

and where $S(t, f)$ is the power with that frequency of that time-frame.

2.3.2 Unselected objective measures

A number of measures which were considered for inclusion in our study (Appendix II) ended up being rejected for various reasons. Some, like the approximate entropy (Pincus, 1991; Rezek & Roberts, 1998; Bhattacharya, 2000; Burioka et al. 2003), had been improved by later versions of what was essentially the same analysis (Richman & Moorman, 2000). So, inclusion of only the latter, improved version was worthwhile.

The redundancy of information among measures of different methodological bases were also considered from a theoretical standpoint, e.g. a NLS with a high percentage of low-energy frames (Shao et al. 2003) would also be identified as being highly 'peaky' in relative amplitude but having few peaks per second. In this way, some measures became redundant to other, more distinguished measures. This lead us to choose measures which could be more directly compared with the majority of the literature but which gave the same theoretical information as other, then redundant, measures. Bandwidth (Shao et al. 2003; Maher & Studniarz, 2012) essentially was an alternative method for describing periodicity, like the HNR, and sounds with a given pitch (Shao et al. 2003) in some frequency range would also have a high RMS in that range.

Other measures, such as state machines (Sasahara & Ikegami, 2004) or multi-scale and multi-resolution entropy techniques (Costa et al. 2003; Costa et al. 2005; Torres & Camero, 2000), required highly specific signal types, making our NLS database unsuitable in these cases. It was also the case that although overall sound intensity had been used in previous studies (Wold et al. 1996; Shao et al. 2003; Maher & Studniarz, 2012), our NLS database had been controlled for this variable already via RMS-normalisation.

2.4 Data analysis

2.4.1 Determining salience of objective measures and independentness of the percepts and accuracy of naming

Since one of our hypotheses of the study was to determine the relationships between objective measures and the subjective percepts and accuracy of naming of NLSs, linear regression was the first utilized technique for data analysis. However, we were first interested in the inter-relationships between the subjective percepts and accuracy of naming, as well as the inter-relationships among the objective measure. This was because it may not be necessary to include two highly and significantly correlated percepts or objective measures, since knowing one allows you to know the same or the opposite of the other.

Linear regressions were therefore conducted in a pair-wise fashion, first for all combinations of subjective percepts and accuracy of naming, and then for all combinations of objective measures within each domain. However, due to temporal and spectral domains being affected differently in different types of hearing loss (Strouse et al., 1998; Probst et al., 2006), we chose to split up the objective measures into these two categories - temporal and spectral - and conducted the subsequent analyses for these categories separately. Significance and correlation coefficients were calculated for all regressions. The resulting regression tables were analysed to make qualitative or quantitative judgements, separately, for which percepts (including accuracy of naming) were most independent, and which objective measures could be chosen as salient measures for inclusion in the subsequent, more complex analyses.

2.4.2 Linear regressions

After determining which of the percepts and accuracy of naming were most independent, and choosing a subset of the original measures to be our considered our salient measures, we conducted linear regressions for all pair-wise combinations between the percepts and accuracy of naming, and the objective measures. Significance and correlation coefficients were again calculated for all regressions, and these relationships were further visualised using scatter-plots with a line-of-best-fit.

2.4.3 NLS database homogeneity

Now knowing the general trends within the data for each of the salient measures and the percepts of accuracy of naming for the NLSs, we tested the homogeneity of the NLS database for their objective measures, as well as their percepts and accuracy of naming. This was because that if the database was insufficiently homogenous, then stronger relationships might be found by removing NLSs from the database which were effectively outliers to the general trend between an objective measure and subjective percept or accuracy of naming. However, to ensure such outlier identifications were not random and had a consistent logic basis, we first identified for each NLS a source category (Appendix I). These categories were experimenter-determined, but allowed us to conduct comparisons between fundamentally dissimilar subsets of NLSs (at least in terms of source). Determination of these differences was done using one-way ANOVAs, where each NLS category formed its own group and were compared for each salient objective measure, as well as the independent percepts and accuracy of naming. ANOVAs were tested for heteroskedasticity using a Brown-Forsythe test and pair-wise differences among the groups were found using a post-hoc Tukey's test. Although the determination of what would be considered 'homogenous' would have to be arbitrary, we decided that we would calculate the number of individual differences (among the groups within all of the ANOVAs combined), and if there were more differences than there was potential for then the database would be sufficiently heterogeneous to consider removing or separating categories of NLSs from the main database.

2.4.4 Multivariate analyses

a) Multiple linear regression and principle component analysis

After this was determined, we moved onto more complex, multivariate analyses, where combinations of objective measures were used to predict or map to perceptual differences or differences in the accuracy of naming for NLSs. The first multivariate analysis technique we used was multiple linear regression to map combinations of salient measures with the percepts and accuracy of naming. Significance and correlation coefficients were calculated for all regressions. We first did this with just two objective measures, and then with all the salient objective measures within a domain.

The strongest linear, multivariate relationships (as determined by their correlation coefficients) for each percept and accuracy of naming were then visualised. Where there were two measures and a percept or accuracy of naming, a bubble plot was used, where the size of the bubbles (each representing a single NLS) were scaled to the percept or accuracy of naming. However, the relationships which included more than two objective measures could not be visualised in two-dimensional plots, making it difficult to appreciate the individual contributions of each of the salient measures in the regressions.

To allow such visualisation, we used a dimensionality reduction technique known as principal component analysis (PCA). PCA is a statistical tool used to reduce the dimensions of a large dataset for exploratory purposes, data visualisation, or subsequent analysis. This is common when a data set can be described in a number of different ways, or described by a number of different variables, and is therefore said to possess many dimensions. It is often not known which dimensions are the most important. Therefore the reduction of the dimensions in the data to fewer dimensions allows use of these dimensions to differentiate sub-groups within the overall data set or to represent a condensed version of the original dataset at the cost of losing information. In PCA, the data variables are largely dependent on one another; in fact, if data were not correlated then PCA would not be suitable for dimensionality reduction. Since our data did correlate to significant extents, thereby fulfilling the requirement that data variables be dependent on each other, we were able to apply the PCA.

b) Agglomerative hierarchical clustering

An alternative way to answer the question as to whether the salient measures influenced or determined differences in the perceptual domain, was to categorise the NLSs on the basis of objective features, derived from combinations of salient measures, ignoring their source of origin. In effect, the process here is akin to placing all the NLSs in a big basket and drawing them out for grouping based on similarities in combinations of their salient measures to see if these groupings correlate with the percepts or with the accuracy of naming. The logic here is that if there were perceptual differences or differences between these objectively-determined categories, then the inputs to that categorisation process - the salient measures - would have led to these differences.

To objectively categorise the NLSs based on their salient measures, a clustering technique to group similar NLSs together was needed and we chose a method known as agglomerative hierarchical clustering (AHC). This method has been used successfully in similar research on NLSs (Ballas, 1993; Gygi et al. 2007) and, given it is an agglomerative (bottom-up) clustering method, we could confidently select clusters at set points of dissimilarity while not affecting the clustering process. This latter advantage makes it distinct, for example, from methods such as k-means clustering, which also has the disadvantage that it is unable to identify density-based clusters and is susceptible to the production of equi-sized clusters which are not fully separate.

The AHC method was used to separate the NLSs into clusters of similar sounds based upon their sets of objective measures. As with the previous analyses, this was done separately for the temporal and spectral salient measures. The resultant clusters were then compared for dissimilarities in their accuracy of naming, complexity, and pleasantness, using one-way ANOVAs. These ANOVAs were tested for heteroskedasticity using a Brown-Forsythe test and pair-wise differences among the groups were found using a post-hoc Tukey's test. If there were significant differences between these clusters for the percepts or accuracy of naming, it would indicate that different unique combinations of objective measures can lead to significantly different accuracies, complexities, and pleasantnesses.

The AHC analyses were performed using the Euclidean distance (shortest theoretical distance) and Ward's method (attempting to provide the least amount of intra-cluster variation in the resultant clusters). The analyses were repeated with varying dissimilarity cut-offs to determine a finite number of clusters. The first cut-off was chosen *a priori* to provide 10 clusters and subsequent cut-offs were chosen to collapse the number of clusters further, generally to roughly halve the number of clusters which were present in the previous cut-off. These subsequent cut-offs were made at dissimilarity levels where there had been relatively few cluster collapses within the whole data set.

However, in these initial AHC analyses we had different numbers of salient temporal measures (n=4) c.f. salient spectral measures (n=3). Although this problem could not properly be subsumed under the term 'curse of dimensionality' (Kriegel et al., 2009; Zimek et al., 2012), it did mean that the clusters in the temporal AHC had to move through a higher number of dimensions and therefore a lot more feature-space to form clusters. This meant, effectively, that the results in these initial AHC

analyses were not necessarily comparable, i.e. although we could say that objective measures led to subjective differences in both domains, we could not say which domain elicited greater perceptual differences.

To equalise the number of dimensions in the AHC analyses we had two options: to add or remove salient measures from one domain, or to reduce the salient measures in one or both domains to equalise the number of dimensions. Since any removal or addition to the salient measures would breach the cut-off rules chosen earlier (despite these rules being arbitrary), we chose to reduce the dimensionality of the salient measures using PCA. To further ensure the AHC analyses could be compared, we chose to reduce both domains to two principle components (PCs), thereby conducting an identical process for both domains' salient measures.

3. Results

3.1 Inter-relationships between percepts and relationship to accuracy of naming NLSs

The provided psychophysical data examined the perception of subjects for NLSs along the dimensions of complexity, pleasantness, and familiarity of the sounds, and as well asking them to name each of the sounds. So we first determined if there were any inter-relationships between the different percepts, and whether any of these aligned with the accuracy of naming.

All three percepts and the accuracy of naming showed significant inter-relationships (Table 1; all Pearson correlations had $p < 0.05$). The correlations were not equal: complexity, familiarity, and accuracy of naming had the most robust inter-relationships, while pleasantness had a weaker relationship with the other percepts and especially with accuracy of naming. It is noteworthy that complexity was inversely correlated with familiarity: the more complex a sound, the less likely it was to be rated as a familiar sound. The accuracy of naming was positively correlated with perceptions of familiarity with the sound, and inversely correlated with perceptions of sound complexity. The first relationship was not surprising, but the inverse relationship to complexity of the NLSs has not, to the best of our knowledge, been described before. Clearly there is a strong set of links here, with complexity being inversely related to familiarity which is positively related to accuracy of naming – and hence accuracy of naming is inversely related to complexity.

Given that familiarity was so highly correlated with accuracy of naming and complexity, familiarity was excluded from subsequent analyses as it provided the least amount of independent variation. It could also be argued that accuracy of naming should also be excluded from further analysis since it appeared to be dependent on complexity. However given the novelty and non-intuitive nature of this relationship (i.e., it is unclear why accuracy of naming should be inversely correlated with complexity), we chose to retain the two percepts of complexity and pleasantness, and the outcome measure of accuracy of naming, in our list of subject perceptions of the NLSs.

Table 1. Correlation matrix (Pearson) of perceptions and identification accuracy.

Variables	Complexity	Pleasantness	Accuracy	Familiarity
Complexity		-0.283	-0.701	-0.807
Pleasantness	-0.283		0.197	0.321
Accuracy	-0.701	0.197		0.735
Familiarity	-0.807	0.321	0.735	

***Bolded** values indicate highly significant relationships ($\alpha=0.001$)

3.2 Identification and selection of salient objective measures for NLSs

A number of objective measures of sounds were applied to the large dataset of NLSs to examine if they could explain the perceptions of complexity and pleasantness, and of the accuracy of naming those sounds. However, given that these objective measures of sounds or of stimulus complexity were derived from an extensive literature review, it is possible that for the dataset of NLSs studied here, some measures may be inter-related (i.e., cross-correlated). The reason why some measures would covary or reveal the same variation for a given NLS is because they may measure or estimate similar or identical features of a sound, e.g. the Higuchi and NLD estimations (see Methods) are both measures of fractal dimension estimation, so their results could be expected to significantly correlate or covary. Similarly, some measures complement each other insofar as measuring opposite things, e.g. for a signal to be harmonic and score highly on the HNR measure, it must have a non-flat spectral profile (i.e., varying powers at different frequencies), which could also give it a low value for the SFM. In cases where measures were significantly correlated with one another, some measures would be redundant and not provide a large amount of additional or unique information. Therefore, the inclusion of similar measures in any analysis would be a confounder that may reduce the explanatory power of other measures.

Thus it was important to first determine if the objective measures selected from the extensive literature search were independent of each other for the set of NLSs to be studied. For this, values were calculated for each of the objective measures for our dataset of NLSs and then determined

whether these values were correlated between the different measures. Since we decided to separate all subsequent analyses into the two domains (see Methods), we examined spectral and temporal measures separately, correlating all the objective spectral measures' values for the NLSs against each other, and then correlating all the objective temporal measures' values for the NLSs against each other. The results of these analyses are shown in Tables 2 (spectral measures) and 3 (temporal measures).

As expected, many measures which were theoretically related or which measured similar qualities of an NLS were significantly correlated with one another, e.g. HNR and SFM ($r=-0.822$); Higuchi and NLD estimates ($r=0.812$). In these instances choosing just one of these inter-related measures would provide nearly the same amount of information about a NLS as choosing both for the subsequent analyses. Although there are many significant relationships among the measures, some of them are quite weak, e.g. Higuchi FD estimate and duration are significantly but weakly correlated ($r=0.059$). Therefore, the choice of measures should include some consideration towards the strength of these relationships. However, any such consideration would have to be made on an entirely arbitrary basis since the 'strength' of a correlation is subjective and relative.

Table 2. Correlation matrix (Pearson) for all temporal measures, showing some significant levels of covariance between some measures.

Variables	HFD	NLDFD	Duration	Peaks / sec	Mean peak	SD of peaks	Sample entropy	Permutation entropy	LZ complexity (average binary)	LZ complexity (modified zone binary)	LZ complexity (differential binary)
HFD		0.812	0.059	0.143	0.202	0.256	-0.193	0.115	0.66	0.361	0.277
NLDFD	0.812		0.009	0.233	0.085	0.165	0.019	0.112	0.697	0.577	0.126
Duration	0.059	0.009		-0.242	0.155	0.206	-0.008	-0.04	0.164	0.055	0.18
Peaks/sec	0.143	0.233	-0.242		-0.568	-0.518	0.261	-0.116	-0.024	0.049	-0.149
Mean peak	0.202	0.085	0.155	-0.568		0.9	-0.691	0.127	0.334	-0.074	0.237
SD of peaks	0.256	0.165	0.206	-0.518	0.9		-0.63	0.204	0.419	0.012	0.237
Sample entropy	-0.193	0.019	-0.008	0.261	-0.691	-0.63		-0.019	-0.225	0.403	-0.242
Permutation entropy	0.115	0.112	-0.04	-0.116	0.127	0.204	-0.019		0.419	0.349	0.305
LZ complexity (average binary)	0.66	0.697	0.164	-0.024	0.334	0.419	-0.225	0.419		0.614	0.604
LZ complexity (modified zone binary)	0.361	0.577	0.055	0.049	-0.074	0.012	0.403	0.349	0.614		0.363
LZ complexity (differential binary)	0.277	0.126	0.18	-0.149	0.237	0.237	-0.242	0.305	0.604	0.363	

Table 3. Correlation matrix (Pearson) for all spectral measures, showing some significant levels of covariance between some measures.

Variables	SFM	SSI	HNR	Mean spectral centroid	250-500Hz RMS	500-1000Hz RMS	1000-2000Hz RMS	2000-4000Hz RMS
SFM		-0.462	-0.822	0.22	0.174	0.057	-0.036	0.096
SSI	-0.462		0.53	0.198	-0.336	-0.197	0.045	0.02
HNR	-0.822	0.53		-0.027	-0.243	-0.104	0.114	-0.031
Mean spectral centroid	0.22	0.198	-0.027		-0.445	-0.45	0.021	0.553
250-500Hz RMS	0.174	-0.336	-0.243	-0.445		0.304	-0.246	-0.469
500-1000Hz RMS	0.057	-0.197	-0.104	-0.45	0.304		-0.066	-0.472
1000-2000Hz RMS	-0.036	0.045	0.114	0.021	-0.246	-0.066		-0.056
2000-4000Hz RMS	0.096	0.02	-0.031	0.553	-0.469	-0.472	-0.056	

*Note that because Tables 2 and 3 are symmetrical about the diagonal, only values below the diagonal are bolded where they fulfil the criterion for being selected as a 'salient measure'.

For our dataset of NLSs, we identified the individual objective measures which accounted for large proportions of the variance in other objective measures within the same domain (i.e., spectral or temporal) – i.e., we identified the measures that correlated highly with each other. The arbitrary cut-off point chosen was $r=\pm 0.45$, i.e., if the correlation between two measures was $\geq \pm 0.45$, only one of the pair would be included in subsequent analyses. Ten cross-relationships with $r \geq \pm 0.45$ were found among the temporal measures and seven such cross-relationships among the spectral measures. Retaining only one from each such cross-relationship reduced our original set of 19 measures to a subset of seven **salient measures** (four temporal and three spectral; listed in Table 4) which, independent of each other, accounted for a large proportion of the total variance among all of the measures.

These **salient measures** were subsequently used to correlate with the percepts and accuracy of naming (Table 4). Mean spectral centroid was the best individual descriptor for both accuracy of naming and complexity. However, the only salient measure to individually correlate with pleasantness in a significant way was a temporal measure, mean peak.

Table 4. Correlation matrix (Pearson) for complexity, pleasantness, and accuracy of naming, showing their individual, pair-wise relationships with each of the salient measures.

Salient measures	Accuracy	Complexity	Pleasantness
NLDFD	-0.254	0.265	-0.057
Mean peak	-0.038	-0.039	0.324
Permutation entropy	-0.031	0.106	-0.034
LZ complexity (differential binary)	-0.11	0.162	-0.048
HNR	0.182	-0.248	0.109
Mean spectral centroid	-0.35	0.293	-0.089
1000-2000Hz RMS	0.076	-0.052	-0.021

***Bolded** values indicate significant relationships ($\alpha=0.05$)

These relationships were also visualised using scatter-plots, separately for all salient measures and accuracy of naming (Figure 1), or for all salient measures and complexity (Figure 2), or for all

salient measures and pleasantness (Figure 3). There were significant correlations for most regressions (see Table 4) but it must be noted that there was also quite large scatter for most regressions. Of these relationships:

- Of all the measures, only the mean peak relative amplitude measure significantly correlated with the perception of pleasantness; however, it did not significantly correlate with the accuracy of naming or complexity;
- Two measures - one temporal (permutation entropy) and one spectral (1000-2000Hz RMS) - did not correlate significantly with any of complexity, pleasantness, or accuracy of naming;
- Whenever a measure was significantly correlated with accuracy of naming, it was also significantly correlated with complexity, except in the case of the temporal measure of LZ complexity (differential binary), which was only significantly correlated with complexity; and
- The measure which best correlated with both accuracy of naming and complexity was the mean spectral centroid.

Figure 2. Linear regressions of all salient measures - temporal and spectral - with complexity. Red lines indicate the line of best fit. High degrees of spread are seen in all regressions. R^2 values and their significance for each regression are shown in Table 4.

Figure 3. Linear regressions of all salient measures - temporal and spectral - with pleasantness. Red lines indicate the line of best fit. High degrees of spread are seen in all regressions. R^2 values and their significance for each regression are shown in Table 4.

3.3 NLSs database homogeneity

The absence of strong correlations between any single salient objective measure and the perceptions and the accuracy of naming could be because we grouped together all our NLSs, spanning a wide variety ranging from machine noises to primate and animal vocalisations. To determine if better correlations might be found for percepts or accuracy of naming of particular sub-types of sounds, we examined the spreads of the accuracy of naming, and the ratings of complexity and of pleasantness (Figure 4), and salient measures (Figures 5 and 6) across different categories of NLSs.

However, before this analysis could be conducted, the issue of what categories to use for allocation of sounds had to be addressed. It was initially considered to get a naïve group of subjects to listen to the NLSs and generate their own categories for allocation of the sounds. However, previous studies (Marcell et al. 2007) using this process have shown that when subjects are allowed to use self-determined categories for NLSs, they use widely varying methods, ranging from acoustic similarity to sound imagery or sound source. This produces widely varying categories, even for the same set of NLSs, and under the same experimental conditions can produce many nonsense or obscure categories. Hence we adopted an experimenter-determined categorisation method using a single categorisation rule as it is likely more reliable, especially if carried out by experienced experimenters, and would produce more consistent categories. Categorisation of our NLSs was made by the experimenters, using the source of the NLS as the sole basis for categorisation, and the allocation of our NLSs to these categories is detailed in Appendix I.

The categories selected were: primate (n=14 sounds), animal (n=38 sounds), tool/machine (n=28), non-animal nature (n=11), human non-vocal (n=6), music (n=21), insect (n=5), explosions/guns (n=10), and uncategorised or other (n=25). (N.B. Some NLSs did not fit well into any of the other categories and remained uncategorised as an “other” group.) Having carried out this classification we could now examine if differences existed among the categories of the NLSs database. If there were, it could mean that there was potential for better fits between the percepts or accuracy of naming and various objective measures.

3.3.1 The spreads of accuracy of naming, complexity, and pleasantness across different categories of NLSs

We examined if there were significant differences among the categories for the perceptions of complexity or pleasantness, or in the accuracy of naming the NLSs. For this, we conducted three one-way ANOVAs - one for each of accuracy of naming, complexity, and pleasantness - across the different categories. Each ANOVA included the nine categories summarised in section 3.3 and listed for each NLS in Appendix I. ANOVAs were tested for heteroskedasticity using a Brown-Forsythe test, and pair-wise differences among the groups were found using a post-hoc Tukey's test.

Significant differences for accuracy of naming, complexity, and pleasantness (Figure 4) among the NLS categories were found, with the greatest number of individual differences found in the case of pleasantness. Heteroskedasticity was noted of the ANOVAs for accuracy of naming and complexity, though this was not found for pleasantness. Notably, the primate NLS category was different to at least one or more other groups in each of the ANOVAs - for accuracy of naming it was different to the “other” group; for complexity it was different to the animal, “other”, and explosions/guns groups; and for pleasantness it was different to the nature and music groups. This suggests that, of all the NLS groups, the primate group was most distinguishable (indeed they were the group which was correctly identified most often; Figure 4) but also was relatively distinct in its characteristics compared to the other groups.

Although there were significant differences (Figure 4) for both accuracy of naming and complexity, only one significant pair-wise difference was found for the former and only four significant pair-wise differences were found for the latter. However, a caveat that should be noted for these ANOVAs for complexity and accuracy of naming is that Brown-Forsythe tests indicated heteroskedasticity. In the case of pleasantness, no heteroskedasticity was found and there were a total of 12 individual significant pair-wise differences among the categories. However, if every category was significantly different to each of the other eight categories, it would be possible to find a total of 48 individual pair-wise differences between the groups. Considering that there were only 12 individual differences of these possible 48, these differences represent only 25% of the total potential differences that there could have been among these groups for pleasantness. Therefore, conducting the

subsequent analyses of the NLSs separately for each of the categories mentioned in section 3.3 would unnecessarily reduce their statistical power.

Although the differences for accuracy of naming, complexity, and pleasantness were of interest and were numerous (especially for pleasantness), ultimately we decided not to separate the categories of NLSs based on these differences. As noted above, the differences represented only $\frac{1}{4}$ of the total possible variation or number of potential pair-wise differences among the categories. Their differences were also not all consistent between the percepts or accuracy of naming - no individual pair-wise difference between the same two groups was found in all three of accuracy of naming, complexity, and pleasantness. Further, these differences were not consistent with the differences found among the categories for the salient measures (Figures 5 and 6).

Having found differences among the categories of NLSs for the percepts and accuracy of naming, we now looked from the perspective of the salient measures to determine if there were objective differences among the categories.

Figure 4. Spreads for the percepts and accuracy are shown here for each of the nine experimenter-determined NLS categories based on sound source. Significant differences between categories were found in each of accuracy, complexity and pleasantness, with the greatest differences found in pleasantness. The ANOVAs for both accuracy and complexity had significantly different standard deviations among their groups, whereas the ANOVA for pleasantness did not.

Significant differences among NLS categories for pleasantness^a

Groups	<i>p</i>
Primate vs. Nature	0.0036
Primate vs. Music	< 0.0001
Animal vs. Music	0.0002
Tool/machine vs. Nature	0.0097
Tool/machine vs. Music	< 0.0001
Nature vs. Human non-voc	0.0161
Nature vs. Insect	0.008
Nature vs. Explosions/guns	0.029
Human non-voc vs. Music	0.0007
Music vs. Insect	0.0004
Music vs. Other	0.0003
Music vs. Explosions/guns	0.0006

3.3.2 The spread of salient measures across different categories of NLSs

Figures 5 and 6 show, for each salient measure, the ranges (spreads) of values for all the NLSs in each category; statistical analysis was conducted, using one-way ANOVAs, to determine if there were differences between categories in these spreads of values for the salient measures. This was done to further test whether there was significant heterogeneity in the NLSs database to warrant the separation or removal of some NLS categories from the main database.

There were some significant differences (Figure 5) among the categories for their spreads of temporal measures. For example, for the salient measure of LZ complexity (differential binary), the spread in the “other” group was different to that in the music and animal groups, and the spread in the tools/machine group was different to that in the animal group. In the same vein, the “other” and music groups also differed in their permutation entropy. Overall, however, the ANOVAs found only four significant inter-group differences of a possible 192 among all the salient temporal measures (48 possible differences per ANOVA given each have nine groups and there are four salient temporal measures).

In the case of differences among the groups for their spreads in the salient spectral measures, the only significant differences found were for the HNR measure (Figure 6); there were no group differences for either the 1000-2000Hz RMS or mean spectral centroid measures.

Given that for the majority of salient measures there were no apparent differences between the different categories, this provides a strong argument for considering the NLSs database as a single group of sounds which have an acceptable homogeneity. Hence from this point on, this will be assumed in this study. It is recognised that there were some perceptual and objective differences between the categories vis-à-vis pleasantness and HNR and for this reason, in relevant subsequent analyses, information on the source of the sound (the category) was retained as supplementary information to assist with describing any trends by sources (see Appendix III).

We thus now turned to the substantive question as to whether there were better fits between the salient objective measures and percepts or accuracy of naming by conducting multivariate analyses.

Figure 5. Spreads for each of the temporal salient measures are shown here for each of the nine experimenter-determined NLS categories based on sound source. Significant differences between categories were found in the cases of permutation entropy and the LZ complexity measure using differential binary. However, the only common dissimilarity of these two measures was between the 'Music' and 'Other' NLS categories.

Figure 6. Spreads for each of the spectral salient measures are shown here for each of the nine experimenter-determined NLS categories based on sound source. Significant differences between categories were found only in the HNR values. Compared to the temporal salient measures which showed some significant differences among the categories, there were many more differences for HNR. This highlighted HNR's importance moving forward.

Significant differences among NLS categories for HNR α

Groups	p
Primate vs. Nature	0.0341
Primate vs. Other	0.0023
Primate vs. Explosions/guns	0.0226
Animal vs. Tool/machine	0.0116
Animal vs. Nature	0.0097
Animal vs. Other	< 0.0001
Animal vs. Explosions/guns	0.0063
Tool/machine vs. Music	0.0001
Nature vs. Music	0.0002
Music vs. Other	< 0.0001
Music vs. Explosions/guns	0.0001

3.4 Multivariate analysis

3.4.1 Multiple linear regression: paired measures

The previous analyses have shown that **single** salient measures were not all significant predictors of the different percepts or of naming accuracy of the different NLSs in our database, and that the amount of explained variance was quite low. We therefore examined if **combinations** of measures could provide better explanatory power, using multiple linear regression for this purpose. In this section we examine the results of using **pairs of measures** to explain each of the of the percepts and the accuracy of naming. In keeping with the distinction made previously between temporal and spectral domains, we only combined salient measures from the same domain (spectral measures or temporal measures).

Table 5. Correlation matrix (R^2) for complexity, pleasantness, and accuracy of naming, showing relationships with pairs of salient measures.

Pairs of salient measures	Accuracy	Complexity	Pleasantness
NLDFD + permutation entropy	0.064	0.076	0.004
NLDFD + mean peak	0.065	0.074	0.112
NLDFD + LZ complexity (differential binary)	0.071	0.087	0.005
Permutation entropy + mean peak	0.002	0.014	0.111
Permutation entropy + LZ complexity (differential binary)	0.012	0.030	0.003
Mean peak + LZ complexity (differential binary)	0.012	0.033	0.121
HNR + mean spectral centroid	0.152	0.144	0.019
HNR + 1000-2000Hz RMS	0.036	0.062	0.013
Mean spectral centroid + 1000-2000Hz RMS	0.130	0.089	0.008

***Bolded** values indicate significant relationships ($\alpha=0.05$)

These analyses are summarised in Table 5 and plotted as bubble plots in Figure 7 for the regressions with the highest explanatory power for each percept and accuracy of naming. Table 5 shows that the spectral combination of HNR and mean spectral centroid provided the largest explanation of both accuracy of naming (Figure 7A) and complexity (Figure 7B) than any other pair-

wise combination. In the case of pleasantness, the temporal combination of the mean peak relative amplitude and LZ complexity (differential binary) measure explained the largest amount of variance (Figure 5).

The HNR measure, which was included in both Figures 7A and 7B, shows for accuracy of naming that smaller bubbles (NLSs which were less frequently identified correctly) are more frequent with lower HNRs. Conversely, in Figure 7B, these NLSs with lower HNRs also were more frequently rated as having high complexity (indicated by the grouping of large bubbles on the lower end of the HNR axis). Mean spectral centroid also appeared to separate NLSs roughly - in the case of accuracy of naming, smaller bubbles (NLSs less accurately identified) were more common on the higher end of the axis, whereas in the case of complexity, the smaller bubbles (NLSs considered not very complex) were grouped on the lower end of the axis.

Figure 7C, which shows the mean peak relative amplitude and LZ complexity (differential binary) measure, showed that larger bubbles (those with a high pleasantness) were situated mainly on the higher end of the mean peak relative amplitude axis. However, there was no clear distinctions present for bubble size along the LZ complexity (differential binary) axis, where there seemed to be a fairly equal representation of pleasant and unpleasant NLSs.

Figure 7. A: Mean spectral centroid and HNR measures bubble plot, where the size of a bubble (representing an individual NLS) is scaled to its accuracy of naming. **B:** Mean spectral centroid and HNR measures bubble plot, where the size of a bubble (representing an individual NLS) is scaled to its complexity. **C:** Mean peak and LZ complexity (differential binary) measures bubble plot, where the size of a bubble (representing an individual NLS) is scaled to its pleasantness.

3.4.2 Multivariate relationships

a) Multiple linear regression

Given the better, but still relatively weak, explanatory power of pairs of objective measures for the percepts and accuracy of naming, we conducted two more regressions, now using all the temporal measures in one regression, and all the spectral measures in another regression. These data are summarised in Table 6. As shown there, although the maximum predictive power for each domain of each percept or accuracy of naming rose marginally, there were no substantial increases (Table 6 c.f. Table 5). This implied that we had reached the limit of our predictive capabilities using just the salient measures to predict the percepts of complexity and pleasantness, as well as identification accuracy.

Table 6. Correlation matrix (R^2) for complexity, pleasantness and identification accuracy, showing their relationships with groups of salient measures.

Salient measures grouped by domain	Accuracy	Complexity	Pleasantness
All salient temporal - <i>NLDFD + permutation entropy + mean peak + LZ</i> <i>complexity (differential binary)</i>	0.071	0.098	0.128
All salient spectral - <i>HNR + mean spectral centroid + 1000-2000Hz RMS</i>	0.157	0.145	0.020
* Bolded values indicate significant relationships ($\alpha=0.05$)			

b) Principal components analysis

Table 6 showed us the total explanatory power when salient measures (temporal or spectral separately) were combined and applied to predict the percepts / accuracy of naming. However, and as discussed earlier (see Methods), these relationships are not able to be visualised in two dimensions easily. Due to this limitation, we reduced the dimensionality of the relationships using PCA.

The outcomes from the PCA showing the most predictive regression for each of the two perceptions and for the accuracy of naming are separately plotted using a biplot (Figure 8). An important note here is that although each panel indicates the salient measures which best described the

percepts or accuracy of naming, the amount of variance represented is less than what exists in the original set. This is due to the nature of the PCA - insofar as it is a dimensionality reduction technique, the total original variance within the variables is distilled into the largest possible portion of that variance for a given number of dimensions. In our case, we have used two dimensions in order to allow two-dimensional visualisation in the form of biplots.

For the inter-relationships of the salient spectral measures with accuracy of naming (Figure 8A), HNR and 1000-2000Hz RMS are relatively unaligned (orthogonal to) accuracy of naming. Conversely, the mean spectral centroid lies on the same plane as the accuracy of naming and is almost exactly opposite in direction. This indicates that of the salient spectral measures which predict accuracy of naming, the mean spectral centroid measure is the most contributory factor, and in this instance - since its direction is opposite to the accuracy of naming - there is a negative correlation between the mean spectral centroid and the accuracy of naming. This matches with our earlier finding that, among the salient spectral measures, mean spectral centroid has the highest correlation with accuracy of naming (Table 4). Figure 8B shows that the 1000-2000Hz RMS measure is nearly orthogonal to complexity, and of HNR and mean spectral centroid, the mean spectral centroid is more closely aligned along the plane of complexity. This again corroborates our earlier finding that, among the salient spectral measures, mean spectral centroid has the highest correlation with complexity (Table 4). In Figure 8C, for pleasantness and the salient temporal measures, mean peak relative amplitude is the primary temporal measure which aligns itself roughly with pleasantness, but all of the other salient temporal measures are relatively unrelated (unaligned with the plane pleasantness lies on). This also agrees with our initial correlations which showed that only mean peak relative amplitude significantly correlated with pleasantness (Table 4).

Figure 8. A: Biplot of PCA for the salient spectral measures which best described accuracy of naming. N.B. Only 62.76% of the variance within these variables is represented. **B:** Biplot of PCA for the salient spectral measures which best described complexity. N.B. Only 62.27% of the variance within these variables is represented. **C:** Biplot of PCA for the salient spectral measures which best described pleasantness. N.B. Only 55.79% of the variance within these variables is represented.

3.4.3 Agglomerative hierarchical clustering

As described in the Methods (see section 2.4.4 b), agglomerative hierarchical clustering (AHC) was conducted as an alternative process to test our hypothesis. This was conducted separately for each domain (Figure 9) and significant perceptual differences were observed in ANOVAs of the resultant clusters for both AHCs (Figures 11 to 16).

For clusters resulting from the spectral AHC, there were significant differences among some clusters for both accuracy of naming (Figure 11) and complexity (Figure 12) – the differences at the 10-cluster level were only found in the case of the accuracy of naming, however the differences at the five- and three-cluster levels included identical clusters for both the accuracy of naming and for complexity. However, these differences seen for accuracy of naming at the 10-cluster level rely on cluster 10 having only two NLSs. Importantly, no differences were found between clusters for pleasantness at any level of clustering (Figure 13). This confirms our previous, simpler analyses that spectral measures are strongly related to both accuracy of naming and complexity, but not to pleasantness (Tables 4, 5 and 6).

The clusters which resulted from the temporal AHC (Figures 14, 15, and 16) showed no differences at the 10- or 4-cluster level. However, at the 2-cluster level there were significant differences for both accuracy of naming and complexity. These differences had, however, much higher p-values ($p=0.0306$ for accuracy of naming and $p=0.0470$; Figures 14 and 15) than those from a comparable cluster level for the spectral AHC, the 3-cluster level (where $p=0.0001$ for accuracy of naming and $p=0.0008$ for complexity; Figures 11 and 12). This ties in with our earlier results showing temporal measures correlate significantly but not as strongly as spectral measures with accuracy of naming and complexity (Tables 4, 5, and 6).

The composition of the resultant clusters, in terms of the source-of-origin-category of the NLSs in each cluster, was also noted and is listed in Appendix III tables A and B.

Figure 9. A: Dendrogram of temporal AHC showing the dissimilarity cut-offs chosen to obtain four and two clusters. The 10-cluster level cluster numbers are indicated on the y-axis. The lines indicate the four-cluster (green) and two-cluster (purple) cut-off with their respective cluster numbers. **B:** Dendrogram of spectral AHC showing the dissimilarity cut-offs chosen to obtain five and three clusters. The 10-cluster level cluster numbers are indicated on the y-axis. The lines indicate the five-cluster (green) and three-cluster (purple) cut-off with their respective cluster numbers.

Figure 11. ANOVAs for resultant clusters of the spectral AHC and their differences for accuracy. Significant differences were found at all three cluster levels, with significance increasing the more the clusters collapsed.

Figure 12. ANOVAs for resultant clusters of the spectral AHC and their differences for complexity. Significant differences were found only at the five- and three- cluster levels, with significance increasing the more the clusters collapsed.

Figure 13. ANOVAs for resultant clusters of the spectral AHC and their differences for pleasantness. No significant differences were found among any of the groups at any cluster level.

Figure 14. ANOVAs for resultant clusters of the temporal AHC and their differences for accuracy. The only significant difference among any of the clusters at any level was at the two-cluster level.

Figure 15. ANOVAs for resultant clusters of the temporal AHC and their differences for complexity. The only significant difference among any of the clusters at any level was at the two-cluster level.

Figure 16. ANOVAs for resultant clusters of the temporal AHC and their differences for pleasantness. No significant differences were found among any of the groups at any cluster level.

3.4.4 Agglomerative hierarchical clustering of principal components

It was expected that, in the previous AHC analyses (see section 3.4.3), there would be significant differences for pleasantness among the clusters for the temporal AHC (Figure 16). This was because the mean peak relative amplitude temporal measure (Table 4), or combinations of temporal measures with mean peak relative amplitude (Tables 5 and 6), were the only salient measures to significantly correlate with pleasantness. Given that the previous temporal AHC analysis did not reveal significant differences in pleasantness (Figure 16), it is possible that the mean peak relative amplitude measure differences are being masked by the other three dimensions of the temporal measures. This is in comparison to the spectral AHC, which only had three salient measures included in total, and where differences were expected - for complexity and accuracy of naming - since there were significant correlations between these included spectral measures in our earlier results (Tables 4, 5, and 6).

We therefore decided to repeat the AHC analyses but both with an equal number of dimensions so they could be more confidently compared. As discussed in the Methods, we could choose to either equalise the number of salient measures in both domains by reducing those in the temporal domain or adding an additional salient measure in the spectral domain. Alternatively we could employ a dimensionality reduction technique as was used earlier in section 3.4.2, where PCA was used to visualise four or five dimensions. Since we have already established that all our salient measures are important, we decided to use PCA to reduce the dimensionality of both sets of salient measures - temporal and spectral. We did this accepting the fact that less of the original variation would be able to be represented in both domains.

PCAs were conducted using all of the temporal measures and, separately, all of the spectral measures. The first two principal components (PCs), which accounted for the greatest proportions of variance within the original sets of salient measures, were then selected as the variables for use in the AHC analyses. In the case of the spectral PCA this was 70.78% of the variance and for the temporal PCA this was 61.37%.

The temporal PC AHC analysis (Figure 17A) found differences for accuracy of naming (Figure 18) and complexity (Figure 19) at different cluster levels. However, these differences were

more significant in the case of accuracy of naming. There were again no significant differences to be found between the clusters for pleasantness (Figure 20). This is possibly due to the fact that only 61.37% of the variance was captured in the PCA and also that of our salient measures, mean peak relative amplitude is the main or sole contributor to pleasantness perception.

Results for the spectral PC AHC analysis (Figure 17B) had similar results as the previous spectral AHC (Figures 11, 12, and 13) - there were differences at multiple cluster levels for accuracy of naming (Figure 21) and complexity (Figure 22), but not for pleasantness (Figure 23). However, in the case of accuracy of naming and complexity, there were no differences at the two-cluster level, and at the 10 and six cluster levels the groups within their ANOVAs had significantly different standard deviations.

The composition of the resultant clusters, in terms of the source-of-origin-category of the NLSs in each cluster, was also noted and is listed in Appendix III tables C and D.

Figure 17. A: Dendrogram of temporal PC AHC showing the dissimilarity cut-offs chosen to obtain six and three clusters. The 10-cluster level cluster numbers are indicated on the y-axis. The lines indicate the six-cluster (green) and three-cluster (purple) cut-off with their respective cluster numbers. **B:** Dendrogram of spectral PC AHC showing the dissimilarity cut-offs chosen to obtain four and two clusters. The 10-cluster level cluster numbers are indicated on the y-axis. The lines indicate the four-cluster (green) and two-cluster (purple) cut-off with their respective cluster numbers.

Figure 18. ANOVAs for resultant clusters of the temporal PC AHC and their differences for accuracy. Differences at each cluster level were observed, with increasing significance as the clusters collapsed.

Figure 19. ANOVAs for resultant clusters of the temporal PC AHC and their differences for complexity. The only significant difference found was at the 3-cluster level, and at both the 10- and 6-cluster level there were no differences.

Figure 20. ANOVAs for resultant clusters of the temporal PC AHC and their differences for pleasantness. No significant differences were found at any cluster level.

Figure 21. ANOVAs for resultant clusters of the spectral PC AHC and their differences for accuracy. Differences were found at both the 10- and 4-cluster levels between multiple pairs of clusters, however at both these stages the standard deviations were significantly different among the clusters.

Figure 22. ANOVAs for resultant clusters of the spectral PC AHC and their differences for complexity. Differences were found at both the 10- and 4-cluster levels between multiple pairs of clusters, however at both these stages the standard deviations were significantly different among the clusters.

Figure 23. ANOVAs for resultant clusters of the spectral PC AHC and their differences for pleasantness. No significant differences were found at any cluster level.

4. Discussion

This study has been successful in making (to our knowledge) the first major attempt at understanding the way humans perceive NLSs for the perception of complexity in relation to the objective features of those sounds. In addition, we have also made a contribution to the ongoing efforts to understand what objectively drives the perception of pleasantness in NLSs, as well as what makes NLSs easy or difficult for humans to accurately name. We have also shown that NLSs in the same category of source-of-origin can have significantly different perceptual and objective properties, as well as different levels of accuracy in naming. We have also provided compelling evidence that, for our data set, various percepts and accuracy of naming are interrelated, e.g., sounds with higher complexities also were rated as not very familiar. Finally for our database of NLSs, various objective measures could carry complementary or even identical information, just in slightly different forms, e.g., sounds with a high HNR also have low values for their SFM.

4.1 Interactions among percepts and accuracy of naming

While there have been many studies probing various perceptual properties of NLSs (Halpern et al., 1986; Cykowicz & Friedman, 1998; Kumar et al., 2008; Reddy et al., 2009; Reuter & Oehler, 2011; Singh, 2011; Kirmse et al., 2012; Lewis et al., 2012; Talkington et al., 2012), few have considered multiple properties in combination (Ballas, 1993; Marcell et al., 2007). Ballas (1993) used the most exhaustive list of perceptual properties, totalling 22 different rating scales, rated by 20 listeners of 41 NLSs. These included dimensions such as 'round-angular' and 'interesting-boring', as well as highly specific questions such as *"How necessary is it to envision this sound in a sequence of sounds in order to identify it?"* and *"How many sounds can you think of that you would place in the same category as this one?"* (Ballas, 1993). This large battery of perceptual reports were condensed into three PCs using PCA to represent 87% of the variance, which is surprising given how many different rating scales were recorded but also telling of the fact that subjective ratings can be highly interrelated or dependent on one another. However, these perceptual and cognitive ratings of the NLSs did not include the accuracy of naming, nor did they include complexity or pleasantness. Ballas

(1993) did, however, include the question "*How familiar does this sound seem to you?*" however this was the fourteenth perceptual question/rating asked of participants (Appendix C of Ballas, 1993) and the methodology allowed participants to replay the sound they were rating as many times as they liked while rating it. This presents a major methodological flaw in that participants may have been growing more and more familiar with the NLSs throughout the rating process. Likewise, this repeating of the stimuli could also affect many of the other perceptual reports; we know that familiarity with a NLS can alter the way it is processed (Cycowicz & Friedman, 1998; Kirmse et al., 2012). The other major study that simultaneously looked at multiple perceptual properties of NLSs was by Marcell et al. (2007), who included the perceptions of complexity, pleasantness, and familiarity, as well as including the accuracy of naming. However, although Marcell et al. (2007) ranked their 22 NLSs for each of these perceptions and the accuracy of naming, they did not attempt to correlate these perceptual ratings.

Given this, no studies (to the best of our knowledge) have attempted to correlate the percepts of complexity, pleasantness, familiarity, and the accuracy of naming. Although our results show that all of these percepts and accuracy of naming are interrelated to various extents for our NLSs database, a dominant relationship was found between the percepts of complexity and familiarity, and for the accuracy of naming. In this relationship, sounds rated as being highly complex are difficult to accurately name and are rated as not familiar, and vice-versa. This leads to two conclusions: (i) a person's auditory experience determines their ability to identify NLSs; and (ii) 'complex' NLSs are rated as such due to a person's lack of familiarity with them, independent of any objective characteristics of the sound's contents.

Unlike the complexity/familiarity/accuracy-of-naming interrelationship, pleasantness was not as well related with the other percepts and especially not with the accuracy of naming. However, of these comparatively weaker relationships, we do tend to find sounds which are familiar are also slightly more pleasant. This could be for one of three reasons: (i) we consciously or unconsciously avoid becoming familiar with sounds which are not pleasant (perhaps by controlling or influencing our auditory environment); (ii) we come to find sounds which we are familiar with 'pleasant' and do

so independent of any objective properties of those sounds; or (iii) we treat unfamiliar sounds as initially unpleasant in order to help avoid or treat carefully potential unknown dangers.

The idea that we would (to some extent) control our auditory environment to select for 'pleasant' sounds is persuasive in the context of the hypothesis that the functional purpose of unpleasantness as an auditory feature is as a means of communicating distress or warnings (Halpern et al., 1986). From an evolutionary standpoint, this hypothesis would suggest that humans naturally or inadvertently select for, control, or influence environments to sound less 'dangerous' or 'distressing'. Thus, an individual with a trait which causes them to avoid unpleasant sounds would also make them more likely to avoid or mitigate the potential dangers or stressors which would be associated with environments where distress or warning calls occurred or were common. Such a trait might also be beneficial in that by responding to an unpleasant sounding environment, i.e. controlling the environment to become less stressful or removing oneself from it, the net stress the individual is exposed to might be lower than if the individual was unresponsive. Being exposed to an overall lowered net stress would also make the individual less susceptible to a variety of disease states and illnesses (Selye, 1955; Cohen et al., 2007) in addition to the direct benefits of survival by avoiding or mitigating the stressor.

However, the basis for this hypothesis (Halpern et al., 1986) relies on the similarities between the properties of unpleasant sounds and macaque monkey warning cries, and these similarities may require retesting due to not being fully reported (Cox, 2008). Further, McDermott & Hauser (2004) showed that cotton-top tamarins (*Saguinus oedipus*), a New World monkey species, had no significant preference for the sound of a three-pronged metal garden tool scraped down a pane of glass (a 'screech' sound which is comparable to fingernails scraping down a blackboard; Halpern et al., 1986) and amplitude-matched white noise (however there was a weak difference and tamarins almost *did* significantly prefer the white-noise control over the screech, $p=0.15$). Humans, comparably, overwhelmingly preferred the white-noise control ($p<0.00001$; McDermott & Hauser, 2004). This difference in preference is in spite of tamarins and humans sharing many similarities in perceptual processing for vocalisations (Ramus et al., 2000; Miller et al., 2001; Miller et al., 2001; Newport et al., 2004). However, McDermott & Hauser (2004) also showed that the tamarins significantly

preferred ($p < 0.01$) looping soundscapes composed of tamarin food chirp sounds c.f. tamarin distress screams. This suggests that unpleasantness might: either (i) be more species-specific than non-specific in terms of objective properties (see section 4.5 for discussion of the objective properties related to pleasantness in humans); or (ii) involve semantic or emotional contents which cannot be fully interpreted by another species. Alternatively still, it could be a combination of both. Considering that humans and New World monkeys last shared a common ancestor 35 million years ago (Schrager & Russo, 2003), species-differentiation of what is considered acoustically unpleasant may have occurred during the intervening time.

Oppositely, we may have also developed an evolutionary advantage by finding some sounds pleasant. By seeking out highly pleasant sounds, such as nature sounds (Shimai et al. 1993; Marcell et al., 2007; Kumar et al., 2008; Alvarsson et al., 2010), an individual would be more likely to stay in a relatively safe environment, e.g. one with running water. Alternatively, we may have learnt to associate safety and survival with sounds found in safe environments. In either case, by finding these 'safe sounds' pleasant, an individual would have an increased chance of survival. In addition, hearing these sounds may further benefit psychological wellbeing or decrease their psychological stress via this 'safe' association. Thus, like the avoidance of unpleasant sounds, this would again decrease an individual's susceptibility to stress-related disease (Cohen et al., 2007). In fact, Alvarsson et al. (2010) show that sounds such as nature sounds (which are consistently rated as highly pleasant; Shimai et al. 1993; Marcell et al., 2007; Kumar et al., 2008; Alvarsson et al., 2010), facilitate recovery from sympathetic activation in humans after experiencing a psychological stressor but that other sounds, such as road traffic sounds (which are often rated as unpleasant; Marcell et al., 2007; Alvarsson et al., 2010), do not facilitate the same recovery. Whether or not this facilitation or absence thereof is primarily due to the objective features of these NLSs or whether higher-order emotional or subjective elements play a role remains to be decisively shown.

4.2 Interactions among objective measures

Many correlations were found among the objective measures in both the temporal and spectral domains. Although some of these relationships were expected and some were not, it is important to emphasise that all of these relationships may only be true for our data set of NLSs and may not be a general rule for all types of NLSs, let alone all types of signals.

4.2.1 Interactions among temporal measures

Although dissimilar in methodology, both fractal dimension estimation techniques - Higuchi and NLD - were highly correlated. This correlation was reported in the original NLD paper (Kallauzi et al., 2009) which applied these analyses to EEG waveforms and not sounds, but we have now extended this similarity in the analysis of NLS waveforms. Interestingly, we have also reported the first significant correlation between a fractal dimension estimate and the LZ complexity measure (average binary and modified zone binary). Although this finding was inadvertent, since both of these complexity measures are ultimately attempting to find self-similarity or repeating aspects within a signal it is reasonable that they should carry some common information. However, since the strength of these correlations were much weaker in the case of the Higuchi FD estimation method c.f. the NLD FD estimation method, it would seem that very slight measure methodology changes impact significantly on this relationship. For example, the NLD FD estimation method correlated with the LZ complexity (modified zone binary) with relatively high strength ($r=0.577$; Table 2) whereas with the same version LZ complexity, the Higuchi FD estimation method was not as well correlated ($r=0.361$; Table 2).

As should be expected among very similar measures or versions of measures, the LZ complexity measures were all well correlated with each other, as were the peaks-related measures. However, it was surprising to find that such a strong, negative relationship existed between mean peak relative amplitude and sample entropy, since they measure completely different things (Richman & Moorman, 2000; Lake et al., 2002 c.f. Gygi et al., 2007). However, this relationship is made understandable when considering the fact that sample entropy ultimately estimates the likelihood of

two identical strings of some length being similar for the next number in the sequence (Richman & Moorman, 2000; Lake et al., 2002), and, in our case, for the next value in the NLS waveform. Where a NLS has a low mean peak relative amplitude, the average peak within that NLS will be closer to the mean relative amplitude of all points. Since it becomes closer to the mean relative amplitude across the whole signal, i.e. 0.5, it becomes more likely that a proceeding point will be similar given the higher frequency of data-points about the mean.

4.2.2 Interactions among spectral measures

The most significant relationship among the spectral measures, the negative relationship between the SFM and HNR, is logical since if a sound is highly harmonic it cannot also be 'flat' in its spectrum. The fact that the SFM and SSI are related explains why they were also well correlated and therefore why there was a relationship between SSI and HNR.

The relationships between the various RMS measures and the mean spectral centroid is also easily explained; if a sound has a relatively high power in the frequency band of 1000-2000Hz, it is unlikely to also have a high power in the lower frequency band of 250-500Hz. Since the mean spectral centroid will rise with increased power at higher frequency bands and fall at lower ones, the relationships between the RMS measures and the mean spectral centroid are similarly predictable.

4.3 Sound source differences

4.3.1 Accuracy of naming of different categories of sound and differences from deaf people

Unlike Inverso & Limb (2010), we found that non-linguistic human vocalisations (classed under 'primates' in our study) were the most accurately identified NLS category and that mechanical sounds (classed under 'tool/machine' in our study) were less well identified, though this difference was not significant. Indeed, the only significant difference observed in our study for accuracy of naming was between the 'primate' and 'other' categories, with primate sounds being significantly ($p=0.0037$) more accurately named.

The differing results could simply be due to the fact that different NLSs were used in both studies and that the categorisation of these NLSs was performed by different experimenters in the Inverso & Limb (2010) study versus our study. However subject differences may also impact on the difference between the two studies since only normal-hearing listeners participated in our study and only experienced cochlear implant (CI) users participated in the Inverso & Limb (2010) study and it is not known if the perception of NLSs is the same between the two categories of people. Inverso & Limb (2010) only used NLSs that were totally accurately identified by normal-hearing listeners and found that the experienced CI users only had about 60-70% accuracy (depending on sound category) in naming the sounds. However, that difference was interpreted as being due to the acoustic characteristics of the sounds (Inverso and Limb, 2010). No consideration was taken as to whether the sounds actually carried the same “meaning” or valency for the CI users, all of whom had been deafened for years, in many cases > 15 years. This may likely be because these participants had lost hearing after the age of 16 years, and so a natural assumption is that, having had quite extensive auditory experience before the onset of deafness, the meaning, saliency, and emotional content of NLSs was the same as that for normal-hearing participants. To the best of our knowledge, this has not been established and it remains a moot point as to whether long-term deafness changes the valency of NLSs.

In this context it is interesting to note that congenitally blind people have been shown to be hypersensitive to hot and cold pain and to rate suprathreshold heat pain as more painful, but show normal responses to non-noxious thermal stimuli (Simani et al., 2013) and this has been suggested to indicate that congenitally blind subjects are more attentive to signals of external threats. Congenitally blind subjects are poorer at taste detection and identification (in contrast to superior performance in auditory, tactile, and olfactory tasks) and this appears to be related to the fact that these subjects do not use external contextual or emotional cues for taste processing but use internal hunger or satiety cues (Gagnon et al., 2013). Thus, congenital deprivation of one sense appears to modify the perception of external events. Whether this is true of only congenital deprivation or applies in long-term deprivation is not known, nor whether it applies in the auditory system, but these findings raise

caveats about assuming that the understanding of the perception of NLSs by normal-hearing subjects can be directly translated to the perception of NLSs by deaf subjects using a cochlear implant.

4.3.2 Differences in the perceptions of different categories of sounds

In the case of complexity perception, no previous study (again, to our knowledge) has reported differences among NLS source categories. Marcell et al. (2007) did report the rank order of 22 NLSs which had different source categories, however did not group the sounds into these categories for comparison to find which category-by-source of NLS was most complex. Notably, for our database of 158 NLSs, primate sounds were significantly less complex than animal sounds, which reinforces the idea that high ratings of 'complexity' are due to a listener being unfamiliar with a sound and that this relationship exists independently of objective qualities, i.e., we would expect human listeners to be more familiar with primate sounds than with non-primate animal sounds. We also found a large number of significant differences in the case of pleasantness. Of the categories we determined for our database, nature and music sounds were rated as the most pleasant, which matches exactly with the results of previous studies (Shimai et al. 1993; Marcell et al., 2007; Kumar et al., 2008; Alvarsson et al., 2010).

In terms of objective differences among the NLS categories, the only salient measures which showed individual differences among the categories were for the LZ complexity measure (differential binary) and the HNR. Although previous studies (Gygi et al., 2007; Lewis et al. 2009; Leaver & Rauschecker, 2010) have noted the importance of harmonicity to NLS classification, none have attempted to study if there were differences among different categories of sounds (experimenter-determined) for HNR. This is also true for LZ complexity. Such differences, as will be discussed later, may influence NLS perceptual differences or identification success.

However, although we did find these significant differences among sound sources for perceptual and objective qualities, there were many groups which were very similar and where no significant differences were observed. Indeed, figures 5 and 6 suggest that sound source does not necessarily segregate NLSs into objectively distinct or coherent groups, i.e., objective features can be similar or identical between different source categories. This partly led to our later categorisation

based on objective measures, i.e., using AHC analyses, which shows that such objective categorising is more revealing of the fundamental make-ups and differences of sounds as it relates to complexity and the accuracy of naming. This allows us to speculate that while the process of identifying a sound's source may rely on objective measures for informational content, it is not purely based on objective features and is somewhat reliant on other information, perhaps such as visual integrative learning or memory. The same is likely true for the perceptions of complexity and pleasantness, i.e., higher-order associations provide context and input into the perception of something as highly complex or pleasant.

4.3.3 Implications for the training of CI users to perceive NLSs

This hypothesis has implications for the identification of NLSs by CI users like those tested by Inverso & Limb (2010). As elaborated above, the lack of success of CI users in identifying NLSs may have less to do with the objective properties of the NLSs and more to do with other attributes, e.g., remembering and associating that sound with an image. Thus if different people (whether CI users or not) tag the same NLSs differently in terms of salience, based on experience or emotional valency, then the ability to recognise the NLSs may depend more on those things. Practically, this may mean that more training is needed for CI and hearing aid users to be able to recognise NLSs than is required to recognise speech, since training for speech is naturally more common in an everyday experience. Additionally, unlike NLSs, speech can include many more - and a broader variety - of visual and contextual cues to aid in this training process. In fact, work on the perceptual learning of spectrally degraded speech and NLSs (Loebach & Pisoni, 2008) shows that speech training does not generalise to NLSs but that NLS training *does* generalise to speech. This suggests that the development of NLS training programs for CI or hearing aid users would be greatly beneficial, giving even more value to the comprehensive characterisation of a NLSs database as has been done in this study.

4.4 Complexity, familiarity and accuracy of naming

Since familiarity so closely matched complexity and, to a lesser extent, the accuracy of naming, any relationship which was true for complexity or accuracy of naming was assumed as existing

oppositely for familiarity. Of these relationships, it was more common for objective measures or combinations thereof to correlate with both accuracy of naming and complexity rather than just one or the other. This gave further weight to the initial relationship found between them, but also showed that they can be influenced or at least correlate with objective measures instead of being purely independent of them and relating only to a person's auditory experience, i.e., familiarity with a NLS.

However, it could also be true that it just so happens we are more familiar or unfamiliar with sounds possessing particular objective feature-sets, e.g., NLSs with a high spectral centroid or a high FD estimate had a negative correlation with accuracy of naming - perhaps it is just that our subjects (who were young, suburban university students) were just less familiar with sounds with a high spectral centroid or high FD estimate. But, then, this begs the question as to why. Indeed, figures 5 and 6 show that there are *no* significant differences among the categories of NLSs for either of FD estimates or the mean spectral centroid. This means that it would have to be by nearly completely chance that our subjects, on average, were less familiar with sounds which so happened to have either high FD estimates or high mean spectral centroids. But instead of by chance, could it be that such sounds were unfamiliar due to the avoidance of such sounds based on the unpleasantness avoidance hypothesis proposed by Halpern et al. (1986)? If it was, we should expect that both of these measures - FD estimation and HNR - would significantly correlate with the percept of pleasantness, whereas we found that neither did. Therefore, the most plausible explanation for why NLSs with high FD estimates or high mean spectral centroids are rated as more complex and are more difficult to identify accurately is due to the objective features themselves as described by these measures.

Against this trend are the effects seen with respect to the HNR. NLSs with a high HNR were easier to accurately identify and are rated as less complex. However, as shown in figure 6, HNR also varies significantly between the NLS categories. In this case, it is possible that sounds with different HNRs have become more or less familiar from experience. Due to the fact that HNR has the second highest correlation of any salient measure to pleasantness (though this relationship is not quite significant), the Halpern et al. (1986) hypothesis would suggest that such differences are due to humans selecting for or controlling their environment to incorporate sounds which are more pleasant (and perhaps more harmonic). However, it could also be said that harmonicity as a feature of sound

has become highly familiar to humans and other primates due to primate vocalisations being highly harmonic. Such familiarity with harmonic sounds might generalise to easier identification of other harmonic, non-primate sounds such as music sounds.

Our later multivariate analyses continued to show that both temporal and spectral measures can be associated significantly with differences in complexity and the accuracy of naming (Tables 5 and 6). By separating the sounds within our database based on these objective measures using AHC analyses, we found that differences arose between the resultant clusters for both complexity and the accuracy of naming. However, the resultant clusters from AHC analyses which included only spectral measures tended to show greater numbers of and more significant differences in complexity and accuracy of naming. As noted in section 3.4.4, this could be due to methodological constraints insofar as the differences either in the number of salient measures between the two domains (as in section 3.4.3), or because of less of the original variance was captured the PCs after dimensionality reduction (as in section 3.4.4).

4.5 Pleasantness

Temporal measures or combinations thereof had the highest explanatory power for ratings of pleasantness, particularly the mean peak relative amplitude of the NLS' waveform. None of the salient spectral measures were significantly correlated - alone or in combination - with pleasantness and clusters created using spectral measures did not show any significant differences in pleasantness. The fact that temporal information becomes highly important in CI users with spectrally degraded stimuli (Fu et al., 2004), and because a decreased level of music appreciation is found in the same individuals (Nimmons et al., 2008; Philips et al., 2012), makes it likely that temporal features are important to the appreciation of sounds. A lower than average temporal resolution might therefore mean that such features are indiscernible to a person with hearing loss, and thus result in a lowered appreciation of music.

However, this conflicts with studies showing that unpleasantness is related to spectral features (Halpern et al., 1986; Cox, 2008; Kumar et al., 2008; Reuter & Oehler, 2011). Nevertheless, perhaps

unpleasantness has different objective drivers than does *pleasantness*, i.e., the presence of an objective feature might make something unpleasant but its absence has a neutral or no effect. This contradiction could also be due to significant methodological differences between our study and previous studies. In the Halpern et al. (1986) study and its replication (Rueter & Oehler, 2011), NLSs were digitally resynthesised via various filtering paradigms. This, unlike our study, possesses a distinct disadvantage in that the sounds were no longer natural or ecologically valid, and ignores the potential for interactions among different combinations of high-power frequencies occurring simultaneously but outside of the same filtering band in a NLS. Kumar et al. (2008) also did not measure any acoustical features directly from the NLSs. Instead, they used an auditory representation of the sound in the primary auditory cortex as described by a highly sophisticated and theoretical computer model (Shamma, 2003). Cox (2008) also did not actually measure any objective qualities of the sounds, and also conducted the study via the Internet, leaving room for potential reporting inaccuracies or technical inconsistencies. Further, none of these studies tested the main temporal measure which we correlated with pleasantness in our study - mean peak relative amplitude. On top of all this, Cox (2008) used only 34 sounds - all of which were pre-determined as unpleasant, i.e., no pleasant sounds were tested for comparison - and only 16 sounds were used in the Halpern et al. (1986) study and an untold, though presumably similarly small number was used in its replication (Rueter & Oehler, 2011). Kumar et al. (2008), who used the most sounds (n=75) still included less than half as many as were used in our study (n=158). Considering all of these differences in methodology, it seems reasonable to accept our results as valid. This is not to say that pleasantness perception definitely has no spectral bases, but rather that it at least has some temporal ones.

4.6 Limitations

Although we have used a large and diverse database of NLSs, there are some important limitations to our study. Firstly, considering that many previous studies (Bruner et al., 1956; Ballas, 1993; Halberstadt & Niedenthal, 1997; Keller & Stevens, 2004; Repp & Knoblich, 2007; Bergman et al., 2009) have noted there to be many other subjective properties, such as emotive and affective

elements, which also play a role in influencing the way a NLS is perceived, it is entirely possible that such elements may have influenced our study without detection. Additionally, the perceived sound source or physical attributes of the sound source also affect human identification and classification of sounds (Dubois, 2000; Genell, 2008) and these informational or contextual factors can be masked in different real-world situations (Ballas & Mullins, 1991; Best et al., 2005; Aravamudhan et al., 2008; Leech et al., 2009; Gygi & Shafiro, 2011). Also, a confounding influence in all of these studies and ours is that familiarity with the signal and repeat exposures appear to alter subjective perceptions (Ballas, 1993; Keller & Stevens, 2004; Giordano et al., 2010), which has been taken advantage of in the development of auditory icons for use in aviation and other human-computer interaction systems (Doll & Folds, 1986; Gaver, 1986; Momtahan et al., 1993). Although our subjects only listened to the NLSs once, we cannot guarantee that they had not heard any of these recordings previously since some were sourced from online multimedia databases, i.e., prior familiarity with any of the sounds used in our database may have been a confound to other perceptual reports. Further, action effects, where a listener's action alters their perception, have also been observed in the reporting of subjective sound features (Repp & Knoblich, 2007) and could have had effects on our subjects which we were unable to detect.

Another major limitation of our study was the fact that we had only 12, fairly young (mean age 19.8 years, SD 0.57) observers to rate and identify the NLSs in our database. It is also likely that they had similar auditory experiences given they attended a suburban university, meaning our results cannot be widely generalised, e.g., a middle-aged farmer or factory working with hearing loss may have a completely different perception of what is 'complex' compared to a suburban student with normal-hearing.

4.7 Conclusions

While recognising the limitations of our study, this project represents a first and highly critical step to the principled subjective and objective characterisation of NLSs in a meaningful way. This allows the creation of a library of NLSs for the evaluation of impairments in the processing of

NLSs when there is a hearing loss or a cognitive impairment. We have shown that a variety of objective measures and percepts (including the accuracy of naming) interact or overlap for the NLSs in our database. This database - and its characterisation - will form a comprehensive and detailed starting point for a number of logical extensions. The obvious practical implication is in the creation of a NLSs hearing test, which, as described in the introduction, may prove highly beneficial in clinical settings by revealing unique, otherwise unknown information. Such information, e.g., if a person finds temporal or spectral information more difficult to process, could aid clinicians in developing more specific treatment, training, or rehabilitation strategies for their patients. Another potential follow-on project which could have beneficial and practical outcomes is the utilization of objective measures characterising NLSs in an imperfect binary algorithm to digitally separate speech in the presence of noise, i.e., to separate speech from a background of NLSs. If possible, such an algorithm could be incorporated into hearing aids or CIs to dramatically counteract one the most commonly complained symptoms of hearing loss - being unable to understand speech in the presence of background noise.

5. References

- Aboy M, Hornero R, Abásolo D & Alvarez D (2006) Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans Biomed Eng.* 53(11):2282-2288.
- Akeroyd MA (2008) Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol.* 47(S2):S53-71.
- Alvarsson JJ, Wiens S & Nilsson ME (2010) Stress Recovery during Exposure to Nature Sound and Environmental Noise. *Int J Environ Res Public Health.* 7(3):1036–1046.
- Aravamudhan R, Lotto AJ & Hawks JW (2008) Perceptual context effects of speech and nonspeech sounds: the role of auditory categories. *J Acoust Soc Am.* 124(3):1695-1703.
- Ballas JA (1993) Common Factors in the Identification of an Assortment of Brief Everyday Sounds. *J Exp Psych: Hum Per Perf.* 19(2):250-267.
- Ballas JA & Mullins RT (1991) Effects of context on the identification of everyday sounds. *Hum Perf.* 4:199–219.
- Ballas JA, Sliwinski MJ & Harding JP (1986) Uncertainty and response time in identifying non-speech sounds. *J Acoust Soc Am.* 79:S47.
- Bandt C & Pompe B (2002) Permutation entropy: a natural complexity measure for time series. *Phys Rev Lett.* 88(17):174102.
- Best V, Ozmeral E, Gallun FJ, Sen K & Shinn-Cunningham BG (2005) Spatial unmasking of birdsong in human listeners: energetic and informational factors. *J Acoust Soc Am.* 118(6):3766-3773.
- Bergman P, Skold A, Vastfjall D & Fransson N (2009) Perceptual and emotional categorization of sound. *J Acoust Soc Am.* 128(6):3156-3167.
- Besser J, Zekveld AA, Kramer SE, Ronnberg J & Festen JM (2012) New Measure of Masked Text Recognition in Relation to Speech-In-Noise Perception and Their Associations With Age and Cognitive Abilities. *J Speech Lang Hear Res.* 55:194-209.

- Bhattacharya J (2000) Complexity analysis of spontaneous EEG. *Acta Neurobiol Exp.* 60:495-501.
- Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proc Ins Phon Sci.* 17:97-110.
- Boersma P (2001) Praat, a system of doing phonetics by computer. *Glott Int.* 5(9):341-345.
- Bruner JS, Goodnow JJ & Austin GA (1956) *A Study of Thinking.* New York: Wiley.
- Bruhn J, Röpcke H & Hoeft A (2000) Approximate entropy as an electroencephalographic measure of anesthetic drug effect during desflurane anesthesia. *Anesthesiology.* 92(3):715-726.
- Büchler M, Allegro S, Launer S & Dillier N (2005) Sound Classification in Hearing Aids Inspired by Auditory Scene Analysis. *EURASIP J Adv Sig Process.* 2005:387845.
- Burioka, N, Cornélissen G, Halberg F, Kaplan DT, Suyama H, Sako T & Shimizu E
(2003) Approximate Entropy of Human Respiratory Movement During Eye-Closed Waking and Different Sleep Stages. *Chest.* 123:80-86.
- Burt JL, Bartolome DS, Burdette DW & Comstock JR (1995) A psychophysiological evaluation of the perceived urgency of auditory warning levels. *Ergo.* 38:2327–2340.
- Casey M (2001) General sound classification and similarity in MPEG-7. *Org Sound* 6:153-164.
- Cohen S, Janicki-Deverts D & Miller GE (2007) Psychological stress and disease. *JAMA.* 298(14):1685-1687.
- Choi YA, Shim HJ, Lee SH, Yoon SW & Joo E-J (2011) Is Cognitive Function in Adults with Hearing Impairment Improved by the Use of Hearing Aids? *Clin Exp Otorhinolaryngol.* 4(2): 72–76.
- Costa M, Peng C-K, Goldberger AL & Hausdorff JM (2003) Multiscale entropy analysis of human gait dynamics. *Physica A.* 330:53-60.
- Costa M, Goldberger AL & Peng CK (2002) Multiscale entropy analysis of physiologic time series. *Phys Rev Lett.* 89:062102.
- Cox TJ (2008) Scraping sounds and disgusting noises. *App Acoust.* 69(12):1195–1204.
- Crandell CC (1991) Individual differences in speech recognition ability: implications for hearing aid selection. *Ear Hear.* 12(s6):100S-108S.

- Cycowicz YM & Friedman D (1998) Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain Cogn.* 36(1):30-51.
- De Lucia M, Tzovara A, Bernasconi F, Spierer L & Murray MM (2012) Auditory perceptual decision-making based on semantic categorization of environmental sounds. *Neuroimage.* 60(3):1704-1715.
- Doehrmann O, Naumer MJ, Volz S, Kaiser J & Altmann CF (2008) Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia.* 46(11):2776-2786.
- Doll TJ & Folds DJ (1986) Auditory signals in military aircraft: Ergonomic principles vs. practice. *App Ergo.* 17:257–264.
- Dubnov (2004) Generalization of Spectral Flatness Measure for Non-Gaussian Linear Processes. *Sig Proc Let.* 11(8):698–701.
- Dubois D (2000) Categories as acts of meaning: The case of categories in olfaction and audition. *Cognit Sci Q.* 1:35–68.
- Edworthy J, Stanton N & Hellier E (1995) Warnings in research and practice: Editorial. *Ergo.* 38:2145–2154.
- Fu QJ, Chinchilla S & Galvin JJ (2004) The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users. *J Assoc Res Otolaryngol.* 5(3):253-260.
- Gagnon L, Kupers R & Ptito M (2013) Reduced taste sensitivity in congenital blindness. *Chem Senses.* 38(6):509-517.
- Gaver WW (1986) Auditory icons: Using sound in computer interfaces. *HCI.* 2:167–177.
- Genell A (2008) Perception of sound and vibration in heavy trucks. Ph.D. thesis. Chalmers University of Technology, Gothenburg.
- Giordano BL, Rocchesso D & McAdams S (2010) Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. *J Exp Psychol Hum Percept Perform.* 36(2):462-476.
- Graham R (1999) Use of auditory icons as emergency warnings: Evaluation within a vehicle collision avoidance application. *Ergo.* 42:1233–1248.

- Green S (1975) Communication by a graded vocal system in Japanese monkeys. In LA Rosenblum (Ed.), *Primate behavior*(Vol. 4). New York: Academic Press.
- Grey JM (1977) Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am.* 61:1270–1277.
- Grey JM & Gordon JW (1978) Perceptual effects of spectral modifications on musical timbres. *J Acoust Soc Am.* 63(5):1493–1500.
- Gunther R, Shapiro B & Wagner P (1994) Complex Systems, complexity measures, grammars and model inferring. *Chaos Sol Frac.* 4:635-651.
- Gygi B (2001) Factors in the identification of environmental sounds. Doctoral dissertation, Indiana University, Bloomington.
- Gygi B, Kidd GR & Watson CS (2004) Spectral-temporal factors in the identification of environmental sounds. *J Acoust Soc Am.* 115(3):1252-65.
- Gygi B, Kidd GR & Watson CS (2007) Similarity and categorization of environmental sounds. *Percept Psych.* 69(6):839-855.
- Gygi B & Shafiro V (2011) The incongruency advantage for environmental sounds presented in natural auditory scenes. *J Exp Psychol Hum Percept Perform.* 37(2):551-565.
- Haas EC & Edworthy J (1996) Designing urgency into auditory warnings using pitch, speed and loudness. *Comp Control Eng J.* 7(4):193-198.
- Haas E & Edworthy J (2006) An introduction to auditory warnings and alarms. In MS Wogalter (Ed), *Handbook of Warnings* (189–198). New Jersey: Lawrence Erlbaum Associates.
- Halberstadt JB & Niedenthal PM (1997) Emotional state and the use of stimulus dimensions in judgement. *J Pers Soc Psychol.* 72:1017–1033.
- Halpern DL, Blake R & Hillenbrand J (1986) Psychoacoustics of a chilling sound. *Percept Psychoacoust.* 39:77–80.
- Hartley RVL (1928) Transmission of Information. *Bell Syst Tech J.* 7(3):535-563.
- Hellier E, Edworthy J & Dennis I (1993) Improving auditory warning design: quantifying and predicting the effects of different warning parameters of perceived urgency. *Hum Fact.* 35:693-706.

- Higuchi T (1988) Approach to an irregular time series on the basis of fractal theory. *Physica D: Nonlin Phenom.* 31:277-283.
- Huang L, Ju F, Zhang E & Cheng J (2003) Real-time Estimation of Depth of Anaesthesia Using the Mutual Information of Electroencephalograms. *Proc Int IEEE EMBS Conf Neu Eng.* 1:328-331.
- Inverso Y & Limb CJ (2010) Cochlear implant-mediated perception of nonlinguistic sounds. *Ear Hear.* 31(4):505-514.
- Jayant NS & Noll P (1984) Digital coding of waveforms: principles and applications to speech and video. New Jersey: Prentice-Hall.
- Jayaram M, Baguley DM & Moffat DA (1992) Speech in noise: a practical test procedure. *J Laryngol Otol.* 106:105-110.
- Johnston JD (1988) Transform coding of audio signals using perceptual noise criteria. *IEEE J Sel Areas Com (J-SAC).* 6(2):314–332.
- Kalauzi A, Bojić T & Rakić L (2009) Extracting complexity waveforms from one-dimensional signals. *Nonlin Biomed Phys.* 3:8.
- Keller P & Stevens C (2004) Meaning from environmental sounds: Types of signal-referent relations and their effect on recognizing auditory icons. *J Exp Psychol Appl.* 10:3–12.
- Khalatur PG, Novikov VV & Khokhlov AR (2003) Conformation-dependent evolution of copolymer sequences. *Phys Rev E Stat Nonlin Soft Matter Phys.* 67(5.1):051901.
- Kidd GR & Watson CS (1999) Sound quality judgements of everyday sounds. *J Acoust Soc Am.* 106(4):2267-2267.
- Kirmse U, Schröger E, Jacobsen T (2012) Familiarity of environmental sounds is used to establish auditory rules. *Neuroreport.* 28;23(5):320-4.
- Kolmogorov A (1968) Three approaches to the quantitative definition of information. *Int J Comp Math.* 2(1-4):157-168.
- Kumar S, Forster HM, Bailey P & Griffiths TD (2008) Mapping unpleasantness of sounds to their auditory representation. *J Acoust Soc Am.* 124(6):3810-3817.

- Kumar S, Von Kriesstein K, Friston K & Griffiths TD (2012) Features vs. Feelings: Dissociable representations of the acoustic features and valence of aversive sounds. *J Neurosci.* 32(41):14184–14192.
- Lake DE, Richman JS, Griffin MP & Moorman JR (2002) Sample entropy analysis of neonatal heart rate variability. *Am J Physiol Regul Integr Comp Physiol.* 283(3):R789-797.
- Leaver AM & Rauschecker (2010) Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *J Neurosci.* 30(22):7604-7612.
- Leech R, Gygi B, Aydelott J & Dick F (2009) Informational factors in identifying environmental sounds in natural auditory scenes. *J Acoust Soc Am.* 126(6):3147-3155.
- Leech R & Saygin AP (2010) Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex. *Brain Lang.* 116(2):83-90.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ & DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci.* 25(21):5148-5158.
- Lewis JW, Talkington WJ, Tallaksen KC & Frum CA (2012) Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. *Front Syst Neurosci.* 6:27.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR & DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. *Cereb Cortex.* 14(9):1008-1021.
- Li X, Logan RJ & Pastore RE (1991) Perception of acoustic source characteristics: Walking sounds. *J Acoust Soc Am.* 90:3036–3049.
- Loebach JL & Pisoni DB (2008) Perceptual learning of spectrally degraded speech and environmental sounds. *J Acoust Soc Am.* 123(2):1126-39.
- Lorenz EN (1963) Deterministic non-periodic flow. *J Atmos Sci.* 20(2):130–141.
- Looi V & Arnephy J (2010) Environmental sound perception of cochlear implant users. *Cochlear Implants Int.* 11(4):203-27.
- Maher RC & Studniarz J (2012) Automatic search and classification of sound sources in long-term surveillance recordings. *AES 46th Int Conf Aud Foren:*14–17.
- Mandelbrot B (1963) The variation of certain speculative prices. *J Bus.* 36(4):394–419.

- Mandelbrot B (1967) How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension. *Sci.* 156(3775):636–638.
- Marcell, Malatanos, Leahy & Comeaux (2007) Identifying, rating, and remembering environmental sound events. *Behav Res Meth.* 39(3):561-569.
- Mathers C, Smith A & Concha M (2000) Global burden of hearing loss in the year 2000. World Health Organisation: Global Burden of Disease 2000.
- Matichuk and Zaïane (1999) Unsupervised Classification of Sound for Multimedia Indexing. *MDM/KDD* 2000:31-36.
- McDermott J & Hauser M (2004) Are consonant intervals music to their ears? Spontaneous acoustic preferences in a nonhuman primate. *Cognition.* 94(2):B11-21.
- Meyer M, Zysset S, von Cramon DY, Alter K (2005) Distinct fMRI responses to laughter, speech, and sounds along the human peri-sylvian cortex. *Cog Brain Res.* 24(2):291-306.
- Miller CT, Dibble E & Hauser MD (2001) Amodal completion of acoustic signals by a nonhuman primate. *Nature Neuroscience* 4:783-784.
- Miller CT, Miller J, Gil-Da-Costa R & Hauser MD (2001) Selective Phonotaxis by Cotton-Top Tamarins (*Saguinus Oedipus*). *Behaviour.* 138(7):811–826.
- Momtahan KL (1990) Mapping of psychoacoustic parameters to the perceived urgency of auditory signals. Masters thesis. Carleton University, Ottawa, Ontario, Canada.
- Momtahan K, Hetu R & Tansley B (1993) Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergo.* 36:1159–1176.
- Newport EL, Hauser MD, Spaepen G & Aslin RN (2004) Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology.* 49(2):85-117.
- Nimmons G, Kang R, Drennan W, Longnion J, Ruffin C, Worman T, Yueh B & Rubinstein J (2008) Clinical assessment of music perception in cochlear implant listeners. *Otol Neurotol.* 29:149-155.
- Philips B, Vinck B, De Vel E, Maes L, D'Haenens W, Keppler H & Dhooge I (2012) Characteristics and determinants of music appreciation in adult CI users. *Eur Arch Otorhinolaryngol.* 269(3):813-821.

- Pincus SM (1991) Approximate entropy as a measure of system complexity. *Proc Natl Acad Sci USA*. 88(6):2297–2301.
- Pollack I & Ficks L (1954) Information of elementary multidimensional auditory displays. *J Acoust Soc Am* 26:155–158.
- Possamai V, Kirk G, Scott A & Skinner D (2012) Speech in noise testing before and after grommet insertion. *J Laryngol Otol*. 126:1010-1015.
- Probst R, Grevers G, Iro H (2006) Basic otorhinolaryngology: a step-by-step learning guide, 2ed. Thieme Medical Publishers, Stuttgart.
- Radhakrishnan N and Gangadhar B (1998) Estimating regularity in epileptic seizure time-series data. *IEEE Eng. Med. Biol. Mag.* 17(3):89–94.
- Raghavendra BS & Dutt DN (2010) Computing fractal dimension of signals using multiresolution box-counting method. *Int J Info Math Sci*. 6(1):50-65.
- Rahman F (2011) Perception of environmental sounds along the complexity, pleasantness and familiarity dimensions. Unpublished PHY 3990 report, Monash University.
- Ramus F, Hauser MD, Miller C, Morris D & Mehler J (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*. 288(5464):349-351.
- Reddy RK, Ramachandra V, Kumar N & Singh NC (2009) Categorization of environmental sounds. *J Bio Cyber*. 100(4):229-306.
- Repp BH & Knoblich G (2007) Action can affect auditory perception. *Psychol Sci*. 18:6–7.
- Rezek I & Roberts SJ (1998) Stochastic Complexity Measures for Physiological Signal Analysis. *IEEE Trans Biomed Eng*. 44(9):1186-1191.
- Richman JS & Moorman JR (2000) Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol*. 278(6):H2039-2049.
- Riedl M, Müller A, Wessel N (2013) Practical considerations of permutation entropy. A tutorial review. *Eur. Phys. J. Special Topics*. 222(2):249-262.
- Rudner M, Foo C, Rönnberg J & Lunner T (2009) Cognition and aided speech recognition in noise: specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand J Psychol*. 50(5):405-418.

- Reuter C & Oehler M (2011) Psychoacoustics of chalkboard squeaking. *J Acoust Soc Am.* 130(4):2545.
- Sasahara K, Bedau M, Husbands P, Watson R & Ikegami T (2004) Song Grammars as Complex Sexual Displays. *AI IX: Proc 9th Int Conf Sim Synth Liv Sys*:194-199.
- Schrager CG & Russo CA (2003) Timing the origin of New World monkeys. *Mol Biol Evol.* 20(10):1620-1625.
- Selye H (1955) Stress and Disease. *Science.* 122(3171):625-631.
- Shafiro V (2008) Identification of Environmental Sounds With Varying Spectral Resolution. *Ear Hear.* 29(3):401-420.
- Shafiro V & Gygi B (2004) How to select stimuli for environmental sound research and where to find them. *Behav Resh Meth: Inst Comp.* 36(4):590-598.
- Shalizi CR, Shalizi KL & Crutchfield JP (2002) An algorithm for pattern discovery in time series. *SFI Working Paper 02-10-060.*
- Shalizi CR (2001) Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata. PhD thesis. Physics Department, University of Wisconsin, Madison.
- Shamma S (2003) Encoding sound timbre in the auditory system. *IETE Journal of Research.* 49:193–205.
- Shannon CE (1948) A Mathematical Theory of Communication. *Bell Syst Tech J.* 27:379–423+623–656.
- Shao X, Xu C & Kankanhalli MS (2003) Applying Neural Network on the Content-Based Audio Classification. *ICICS-PCM: 3C2.9.*
- Shibayama H (2006) The complexity of environmental sound as a function of seasonal variation. *J Acoust Soc Am.* 120(5):3240.
- Shimai S, Fukuda K & Terasaki M (1993) Pleasantness-unpleasantness of environmental sounds and gender difference in evaluation. *Percept Mot Skills.* 76(2):635-640.
- Singh NC (2011) Measuring the ‘complexity’ of sound. *Pramana J Phys.* 77(5):811-816.
- Slimani H, Danti S, Ricciardi E, Pietrini P, Ptito M & Kupers R (2013) Hypersensitivity to pain in congenital blindness. *Pain.* 154(10):1973-1978.

- Solomonoff RJ (1964) A Formal Theory of Inductive Inference. *Info Control*. 7(1)1–22+(2):224–254.
- Spasic S, Kalauzi A, Culic M, Grbic G & Martac L (2005) Estimation of parameter kmax in fractal analysis of rat brain activity. *Annal NY Acad Sci*. 1048:427-429.
- Strouse A, Ashmead DH, Ohde RN, Grantham DW (1998) Temporal processing in the aging auditory system. *Journal of the Acoustical Society of America*. 104(4):2385-2399.
- Szczepanski J, Amigo JM, Wajnryb E & Sanchez-Vives MV (2003) Application of Lempel–Ziv complexity to the analysis of neural discharges. *Net: Comp Neuro Sys*. 14:335–350.
- Talkington WJ, Rapuano KM, Hitt LA, Frum CA & Lewis JW (2012) Humans mimicking animals: a cortical hierarchy for human vocal communication sounds. *J Neuro*. 32(23):8084-8093.
- Tchernichovski O, Nottebohm F, Ho CE, Pesaran B & Mitra PP (2000) A procedure for an automated measurement of song similarity. *An Behav*. 59(6):1167–1176.
- Torres ME & Gamero LG (2000) Relative Complexity Changes in Time Series using Information Measures. *Physica A*. 286(3-4):457-473.
- Warren WH & Verbrugge RR (1984) Auditory perception of breaking and bouncing events: A case study in ecological acoustics. *J Exp Psychol: Hum Lear Mem*. 10:704–712.
- Watanabe TA, Cellucci CJ, Kohegyi E, Bashore TR, Josiassen RC, Greenbaun NN & Rapp PE (2003) The algorithmic complexity of multichannel EEGs is sensitive to changes in behavior. *Psychophys*. 40(1):77-97.
- WHO (1991) Report of the Informal Working Group On Prevention Of Deafness And Hearing Impairment Programme Planning. Geneva, 1991.
- Wilson RH (2011) Clinical Experience with the Words-in-Noise Test on 3430 Veterans: Comparisons with Pure-Tone Thresholds and Word Recognition in Quiet. *J Am Acad Audiol*. 22:405-423.
- Wilson RH, Trivette CP, Williams DA & Watts KL (2012) The Effects of Energetic and Informational Masking on the Words-in-Noise Test (WIN). *J Am Acad Audiol*. 23:522-533.
- Wold E, Blum T, Keislar D & Wheaton J (1996) Content-based classification, search and retrieval of audio. *IEEE Multimedia Mag*. 3:27-36.
- Wolfe D (2002) Speech and music, acoustics and coding, and what music might be 'for'. *Proc 7th Int Conf Mus Perc Cog*:10-13.

- Wu X & Xu J (2001) Complexity and brain function. *Acta Biophysica Sinica*. 7:103–106.
- Xu J, Liu R & Yang QF (1997) Information transformation in human cerebral cortex. *Physica D*. 106:363–374.
- Xue J, Wichern G, Thornburg H & Spanias A (2008) Fast query by example of environmental sounds via robust and efficient cluster-based indexing. *IEEE-ICASSP 2008*:5-8.
- Yumoto E & Gould WJ (1982) Harmonics-to-noise ratio as an index of the degree of hoarseness. *J Acoust Soc Am*. 71(6):1544-1550.
- Zald DH & Pardo JV (2002) The neural correlates of aversive auditory stimulation. *Neuroimage*. 16(3.1):746-753.
- Zanin M, Zunino L, Rosso OA & Papo D (2012) Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy*. 14(8):1553-1577.
- Ziv J & Lempel A (1978) Compression of Individual Sequences via Variable-Rate Coding. *IEEE Trans Info Theory*. 24(5):530-536.
- Zhang XH, Zhu YS & Wang ZM (2000) Complexity measure and complexity rate information based detection of ventricular tachycardia and fibrillation. *Med. Biol. Eng. Comput*. 38(5):553–557.
- Zhang XS, Roy RJ & Jensen EW (2001) EEG complexity as a measure of depth of anesthesia for patients. *IEEE Trans. Biomed. Eng*. 48(12):1424–1433.

Appendix

Appendix I : List of sounds and assigned categories

	<i>Percentage</i>	
	<i>of data set</i>	<i>Number of sounds</i>
Primate (#1)	8.86	14
Animal (#2)	24.05	38
Tool/machine (#3)	17.72	28
Nature (#4)	6.96	11
Human non-voc (#5)	3.80	6
Music (#6)	13.29	21
Insect (#7)	3.16	5
Other (#8)	15.82	25
Explosions/guns (#9)	6.33	10
	<i>Total</i>	<i>158</i>

<i>File name</i>	<i>Category</i>		
		BOWLING.WAV	#8
Airplane.wav	#3	BRUSHTEE.WAV	#8
BABYCRY.WAV	#1	burnout.wav	#3
barnyard.wav	#2	BURP.WAV	#5
Baskball.wav	#8	CAMERA.WAV	#3
bee.wav	#7	CANCRUSH.WAV	#8
beeswarm.wav	#7	CARCRASH.WAV	#9
bikstart.wav	#3	CARHORN.WAV	#6
bird.wav	#2	Cashreg.wav	#3
Birds.wav	#2	Cat.wav	#2
BLOWNOSE.WAV	#5	catmeow.wav	#2
Boeing-707.wav	#3	catpurr2.wav	#2

catscream.wav	#2	Drums.wav	#6
catyell.wav	#2	duck.wav	#2
Chewing.wav	#5	elephant.wav	#2
Chickens.wav	#2	explosion-01.wav	#9
Chicough.wav	#1	Explsn.wav	#9
chimp.wav	#1	fart-02.wav	#5
Chrbell.wav	#6	firecrackers-01.wav	#9
Clapping.wav	#5	FIRECRAK.WAV	#9
Clearthr.wav	#1	firetrck.wav	#3
coin-01.wav	#8	Flute.wav	#6
Coindrop.wav	#8	Frog.wav	#2
COW.WAV	#2	frog2.wav	#2
crankit.wav	#3	Fryfood.wav	#8
Crickets.wav	#7	GARGLING.WAV	#5
Crow.wav	#2	Glassbrk.wav	#8
crow2.wav	#2	Gong.wav	#6
crunch.wav	#8	GUITAR.WAV	#6
CUCKOOCL.WAV	#2	Gunshots.wav	#9
Cymbals.wav	#6	hairdryer-02.wav	#3
dishes-break-01.wav	#8	Hammerng.wav	#3
Dogbark.wav	#2	HARMONCA.WAV	#6
doggrowl.wav	#2	HARP.WAV	#6
dolphin.wav	#2	hawk.wav	#2
donkey.wav	#2	HELICPTR.WAV	#3
DOORBELL.WAV	#6	horngoby.wav	#6
Doorclos.wav	#8	Horsegal.wav	#8
Dropice.wav	#8	horsegallop.wav	#8

horseneigh.wav	#2	03.wav	
hothonda.wav	#3	PIANO.WAV	#6
hurricane.wav	#4	pig.wav	#2
hurtdog.wav	#2	PINBALL.WAV	#3
JACKHAMM.WAV	#3	Pingpong.wav	#8
jet-start-01.wav	#3	Pourwatr.wav	#4
junglebirds.wav	#2	RAIN.WAV	#4
kiss-01.wav	#8	rain-roof-01.wav	#4
Knocking.wav	#8	Rattlsna.wav	#2
Laughing.wav	#1	rooster.wav	#2
laugh-man-03.wav	#1	running-water-	
LAWNMOWR.WAV	#3	01.wav	#4
lion.wav	#2	Sawing.wav	#3
locust.wav	#7	SAX.WAV	#6
MACHGUN.WAV	#9	SCREAM.WAV	#1
machinegun-01.wav	#9	seagull.wav	#2
miata-hn.wav	#3	Seal.wav	#2
mice.wav	#2	sheep.wav	#2
monkey.wav	#1	shotgun-01.wav	#9
monkeyscream.wav	#1	Shuflcrd.wav	#8
MOSQUITO.WAV	#7	skyrocket-01.wav	#9
Motorcyc.wav	#3	SNEEZE.WAV	#1
Ocean.wav	#4	snore-02.wav	#1
ORGAN.WAV	#6	SNORING.WAV	#1
OWL.WAV	#2	SONAR.WAV	#6
panther.wav	#2	sparrowclose.wav	#2
pencil-sharpen-	#3	STAPLER.WAV	#3

Swords.wav	#8	VIOLIN.WAV	#6
Teakettl.wav	#8	Watrbubb.wav	#4
Tearpapr.wav	#8	Watdrai.wav	#4
TELEPHON.WAV	#6	Watdrip.wav	#4
THUNDER.WAV	#4	Whip.wav	#8
tirespin.wav	#3	Whistle.wav	#6
Toilet.wav	#8	WHISTLNG.WAV	#1
TRAIN.WAV	#3	WIND.WAV	#4
trckstrt.wav	#3	windchimes-01.wav	#6
TRUCK.WAV	#3	Wndchime.wav	#6
TRUMPET.WAV	#6	Wolf.wav	#2
turkey.wav	#2	Woodpckr.wav	#3
Turnpage.wav	#8	YAWN.WAV	#1
Typewrtr.wav	#3	ZIPPER.WAV	#3
VELCRO.WAV	#8		

Appendix II : Considered temporal and spectral measures of complex waveforms

<i>Measure</i>	<i>Description</i>	<i>Used or defined in</i>
<i>Temporal-based measures</i>		
Sample entropy (SE)	A modified version of Approximate Entropy based on information theory which quantifies the probability of similar sequences of data-points existing in the waveform.	Richman & Moorman, 2000; Lake et al. 2002
Permutation entropy (PE)	Based on chaos theory, the randomness of the distribution of possible permutations which exist as sections of the waveform is calculated.	Bandt & Pompe, 2002
Lempel-Ziv (LZ) algorithmic complexity	Attempts to condense the waveform into the shortest possible binary string by enumerating repeated information within the original waveform. Complexity is then quantified as the difference between the original length of the waveform and the binary string.	Ziv & Lempel, 1978; Khalatur et al. 2003; Szczepanski et al. 2003; Watanabe et al. 2003; Huang et al. 2003
Duration	The duration of the sound in seconds.	Ballas, 1993; Gygi et al. 2007; Maher & Studniarz, 2012
Peaks-related measures	Defining a peak as any point in the waveform which is >80% of the waveform's range, we can measure: • Peaks/second - the average number of peaks per second; • Mean peak - the mean amplitude (as a point value) for all peaks; and • Peak variability - the standard deviation of all peaks.	Gygi et al. 2007
Fractal dimension (FD) estimation	A mathematical description of a shape as taking up some portion of theoretical space beyond its topological dimension. Because this measure can only be estimated, two unique algorithms will be used in our study: the Higuchi and Normalised Length Density.	Mandlebrot, 1967; Higuchi, 1988; Kalauzi et al. 2009; Raghavendra & Dutt, 2010; Spasic et al. 2005; Shibayama, 2006
Approximate	Based on information theory and quantifies the probability	Pincus, 1991;

entropy	of similar sequences of data-points existing in the waveform.	Rezek & Roberts, 1998; Bhattacharya, 2000; Burioka et al. 2003
Hidden Markov model states	Attempts to find the minimum number of hidden states required to accurately describe and reproduce, in an algorithmic sense, the waveform.	Gunther, 1994; Shalizi et al. 2001; Shalizi et al. 2002; Buchler et al. 2005
State machine	Assumes a sequential or patterned nature in the waveform and measures complexity as the ratio between the number of possible states and number of possible states from those states.	Sasahara & Ikegami, 2004
Multi-scale and multi-resolution techniques	Better approximates information-theory-based estimates of complexity in very long signals.	Costa et al. 2003; Costa et al. 2005; Torres & Camero, 2000
Percentage of low energy frames	Identifies frames in the waveform which have a low energy and expresses this as a percentage of total frames.	Shao et al. 2003
<i>Spectral-based measures</i>		
Mean spectral centroid	Characterises a sounds' spectrum by its weighted mean of the frequencies it contains and has been shown to correlate with the percept of the 'brightness' of a sound.	Grey & Grodon, 1978; Gygi et al. 2007; Maher & Studniarz, 2012; Shao et al. 2003
Root mean square (RMS) of frequency bands	Provides non-self-relative information about the amount of power at particular frequency ranges in a given sound.	Gygi et al. 2007
Harmonics-to-noise ratio (HNR)	Amount of power in the signal which is harmonic relative to the amount of power in the remainder of the signal (noise).	Yumoto & Gould, 1982; Leaver & Rauschecker, 2010; Boersma, 1993; Lewis et al. 2009
Spectral flatness measure (SFM)	Uniformity of the signal within the power spectrum.	Jayant & Noll, 1984; Boersma,

		2001
Spectral structure variability (SSV) or index (SSI)	How much the <i>SFM</i> varies over time, essentially measuring the periodicity of the signal.	Singh, 2011; Reddy et al. 2009; Talkington et al. 2012
Intensity	The root mean square for a sound.	Maher & Studniarz, 2012; Shao et al. 2003; Wold et al. 1996
Pitch	Fundamental frequency of the sound.	Shao et al. 2003
Bandwidth	Describes the sound as existing in a certain spectral range with the extremes of the axis being tonal at one end and broadband noise at the other.	Maher & Studniarz, 2012; Shao et al. 2003

Appendix III : Category make-up of resultant clusters for AHC analyses

A. Cluster makeup at the 10-, 5- and 3-cluster levels of the spectral AHC. For each cluster, a percentage of the total NLSs of each category is listed to provide an indication of its makeup and also to show the spread of individual categories among the clusters at each level.

<i>10-cluster level</i>	1	2	3	4	5	6	7	8	9	10
Primate	0.00%	28.57%	28.57%	14.29%	7.14%	7.14%	0.00%	7.14%	7.14%	0.00%
Animal	7.89%	15.79%	5.26%	18.42%	5.26%	13.16%	18.42%	5.26%	7.89%	2.63%
Tool/machine	14.29%	21.43%	7.14%	21.43%	14.29%	10.71%	3.57%	7.14%	0.00%	0.00%
Nature	9.09%	9.09%	0.00%	27.27%	18.18%	9.09%	9.09%	18.18%	0.00%	0.00%
Human non-voc	16.67%	33.33%	0.00%	16.67%	16.67%	0.00%	0.00%	16.67%	0.00%	0.00%
Music	4.76%	28.57%	14.29%	14.29%	9.52%	14.29%	0.00%	14.29%	0.00%	0.00%
Insect	40.00%	20.00%	0.00%	20.00%	0.00%	0.00%	0.00%	0.00%	20.00%	0.00%
Other	8.00%	8.00%	12.00%	28.00%	8.00%	8.00%	16.00%	4.00%	4.00%	4.00%
Explosions/guns	10.00%	20.00%	0.00%	0.00%	10.00%	30.00%	10.00%	0.00%	20.00%	0.00%
<i>5-cluster level</i>	1	2	3	4	5	<i>3-cluster level</i>				
Primate	0.00%	35.71%	42.86%	14.29%	7.14%	Primate	7.14%	50.00%	42.86%	

Animal	26.32%	28.95%	23.68%	10.53%	10.53%	Animal	36.84%	39.47%	23.68%
Tool/machine	17.86%	28.57%	28.57%	21.43%	0.00%	Tool/machine	17.86%	53.57%	28.57%
Nature	18.18%	18.18%	27.27%	36.36%	0.00%	Nature	18.18%	54.55%	27.27%
Human non-voc	16.67%	33.33%	16.67%	33.33%	0.00%	Human non-voc	16.67%	66.67%	16.67%
Music	4.76%	42.86%	28.57%	23.81%	0.00%	Music	4.76%	66.67%	28.57%
Insect	40.00%	20.00%	20.00%	0.00%	20.00%	Insect	60.00%	20.00%	20.00%
Other	24.00%	16.00%	40.00%	12.00%	8.00%	Other	32.00%	28.00%	40.00%
Explosions/guns	20.00%	50.00%	0.00%	10.00%	20.00%	Explosions/guns	40.00%	60.00%	0.00%

B. Cluster makeup at the 10-, 4- and 2-cluster levels of the temporal AHC. For each cluster, a percentage of the total NLSs of each category is listed to provide an indication of its makeup and also to show the spread of individual categories among the clusters at each level.

<u>10-cluster level</u>	1	2	3	4	5	6	7	8	9	10
Primate	7.14%	7.14%	0.00%	14.29%	0.00%	14.29%	28.57%	28.57%	0.00%	0.00%
Animal	2.63%	15.79%	2.63%	7.89%	26.32%	7.89%	18.42%	15.79%	2.63%	0.00%
Tool/machine	17.86%	17.86%	14.29%	17.86%	3.57%	17.86%	7.14%	0.00%	3.57%	0.00%
Nature	9.09%	0.00%	18.18%	18.18%	9.09%	0.00%	27.27%	0.00%	18.18%	0.00%
Human non-voc	16.67%	16.67%	0.00%	33.33%	0.00%	16.67%	16.67%	0.00%	0.00%	0.00%
Music	19.05%	14.29%	4.76%	0.00%	9.52%	14.29%	9.52%	0.00%	9.52%	19.05%
Insect	20.00%	0.00%	0.00%	20.00%	20.00%	20.00%	0.00%	0.00%	0.00%	20.00%
Other	20.00%	12.00%	24.00%	20.00%	0.00%	12.00%	8.00%	4.00%	0.00%	0.00%
Explosions/guns	0.00%	0.00%	10.00%	30.00%	0.00%	20.00%	30.00%	10.00%	0.00%	0.00%
<u>4-cluster level</u>	1	2	3	4						
Primate	7.14%	7.14%	57.14%	28.57%	<u>2-cluster level</u>	1	2			
Animal	5.26%	42.11%	34.21%	18.42%	Primate	64.29%	35.71%			
Tool/machine	32.14%	21.43%	42.86%	3.57%	Animal	39.47%	60.53%			
					Tool/machine	75.00%	25.00%			

Nature	27.27%	9.09%	45.45%	18.18%	Nature	72.73%	27.27%
Human non-voc	16.67%	16.67%	66.67%	0.00%	Human non-voc	83.33%	16.67%
Music	23.81%	23.81%	23.81%	28.57%	Music	47.62%	52.38%
Insect	20.00%	20.00%	40.00%	20.00%	Insect	60.00%	40.00%
Other	44.00%	12.00%	40.00%	4.00%	Other	84.00%	16.00%
Explosions/guns	10.00%	0.00%	80.00%	10.00%	Explosions/guns	90.00%	10.00%

C. Cluster makeup at the 10-, 6- and 3-cluster levels of the temporal PC AHC. For each cluster, a percentage of the total NLSs of each category is listed to provide an indication of its makeup and also to show the spread of individual categories among the clusters at each level.

<i>10-cluster level</i>	1	2	3	4	5	6	7	8	9	10
Primate	0.00%	21.43%	7.14%	21.43%	7.14%	14.29%	7.14%	7.14%	0.00%	14.29%
Animal	7.89%	10.53%	5.26%	15.79%	36.84%	2.63%	2.63%	5.26%	5.26%	7.89%
Tool/machine	10.71%	25.00%	17.86%	3.57%	7.14%	7.14%	10.71%	10.71%	3.57%	3.57%
Nature	18.18%	0.00%	18.18%	9.09%	9.09%	18.18%	0.00%	27.27%	0.00%	0.00%
Human non-voc	0.00%	0.00%	16.67%	33.33%	0.00%	16.67%	0.00%	33.33%	0.00%	0.00%

Music	4.76%	33.33%	14.29%	14.29%	4.76%	0.00%	0.00%	9.52%	4.76%	14.29%
Insect	20.00%	20.00%	0.00%	20.00%	40.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other	20.00%	16.00%	16.00%	8.00%	16.00%	12.00%	8.00%	4.00%	0.00%	0.00%
Explosions/guns	10.00%	50.00%	0.00%	0.00%	20.00%	0.00%	0.00%	10.00%	10.00%	0.00%
<i>6-cluster level</i>	1	2	3	4	5	6				
Primate	0.00%	28.57%	21.43%	21.43%	14.29%	14.29%				
Animal	7.89%	15.79%	7.89%	21.05%	39.47%	7.89%				
Tool/machine	10.71%	35.71%	25.00%	7.14%	17.86%	3.57%				
Nature	18.18%	27.27%	36.36%	9.09%	9.09%	0.00%				
Human non-voc	0.00%	33.33%	33.33%	33.33%	0.00%	0.00%				
Music	4.76%	42.86%	14.29%	19.05%	4.76%	14.29%				
Insect	20.00%	20.00%	0.00%	20.00%	40.00%	0.00%				
Other	20.00%	20.00%	28.00%	8.00%	24.00%	0.00%				
Explosions/guns	10.00%	60.00%	0.00%	10.00%	20.00%	0.00%				
<i>3-cluster level</i>	1	2	3				1 (cont'd)	2 (cont'd)	3 (cont'd)	
Primate	14.29%	50.00%	35.71%	Other			44.00%	48.00%	8.00%	

Animal	47.37%	23.68%	28.95%	Explosions/guns	30.00%	60.00%	10.00%
Tool/machine	28.57%	60.71%	10.71%				
Nature	27.27%	63.64%	9.09%				
Human non-voc	0.00%	66.67%	33.33%				
Music	9.52%	57.14%	33.33%				
Insect	60.00%	20.00%	20.00%				

D. Cluster makeup at the 10-, 4- and 2-cluster levels of the spectral PC AHC. For each cluster, a percentage of the total NLSs of each category is listed to provide an indication of its makeup and also to show the spread of individual categories among the clusters at each level.

<i>10-cluster level</i>	1	2	3	4	5	6	7	8	9	10
Primate	0.00%	28.57%	21.43%	7.14%	7.14%	21.43%	0.00%	7.14%	14.29%	0.00%
Animal	5.26%	15.79%	15.79%	10.53%	15.79%	10.53%	15.79%	5.26%	2.63%	2.63%
Tool/machine	7.14%	3.57%	28.57%	17.86%	10.71%	17.86%	0.00%	10.71%	0.00%	0.00%
Nature	0.00%	0.00%	9.09%	27.27%	18.18%	27.27%	0.00%	18.18%	0.00%	0.00%
Human non-voc	0.00%	0.00%	50.00%	0.00%	0.00%	33.33%	0.00%	16.67%	0.00%	0.00%
Music	0.00%	23.81%	28.57%	9.52%	4.76%	0.00%	4.76%	9.52%	19.05%	0.00%

Insect	20.00%	40.00%	0.00%	0.00%	20.00%	20.00%	0.00%	0.00%	0.00%	0.00%
Other	24.00%	4.00%	16.00%	12.00%	0.00%	40.00%	0.00%	0.00%	0.00%	4.00%
Explosions/guns	20.00%	0.00%	10.00%	50.00%	10.00%	10.00%	0.00%	0.00%	0.00%	0.00%
<i>4-cluster level</i>	1	2	3	4						
Primate	0.00%	7.14%	57.14%	28.57%						
Animal	5.26%	42.11%	34.21%	18.42%						
Tool/machine	32.14%	21.43%	42.86%	3.57%						
Nature	27.27%	9.09%	45.45%	18.18%						
Human non-voc	16.67%	16.67%	66.67%	0.00%						
Music	23.81%	23.81%	23.81%	28.57%						
Insect	20.00%	20.00%	40.00%	20.00%						
Other	44.00%	12.00%	40.00%	4.00%						
Explosions/guns	10.00%	0.00%	80.00%	10.00%						
					<i>2-cluster level</i>	1	2			
					Primate	50.00%	50.00%			
					Animal	50.00%	50.00%			
					Tool/machine	85.71%	14.29%			
					Nature	81.82%	18.18%			
					Human non-voc	100.00%	0.00%			
					Music	47.62%	52.38%			
					Insect	40.00%	60.00%			
					Other	96.00%	4.00%			
					Explosions/guns	90.00%	10.00%			