A Pragmatic and Musically Pleasing Production System for Sonic Events

Marc Conrad Tim French Marcia Gibson
University of Luton, Park Square, United Kingdom
Marc.Conrad@luton.ac.uk; Tim.French@luton.ac.uk; Marcia.Gibson@luton.ac.uk

Abstract

We describe a novel application for sonic events namely their generation via mathematical functions implemented on a universal all purpose Java platform. Their design is driven by a set of requirements that arise in recognition-based authentication systems. We show that our approach has potential advantages as compared with traditional alphanumeric and other password systems. Our intention is to demonstrate that by leveraging familiar musical dimension and aesthetics human memorability, pleasure and pragmatics are enhanced. We demonstrate and briefly discuss one exemplar generative approach that has been specifically designed in order to fulfill the requirements implied by authentication systems. It is hoped that this work serves to stimulate debate and further activity in the field of computer generated sonics..

1. Introduction

Music and the computer have quite a lengthy history of association, as with other humanistic disciplines. Familiar inter-developmental and synergistic uses include computer-aided composition, computer-aided musical stylistic analysis, computer-aided studies of performance and acoustic analysis, as well as the vast arena that today broadly comprises MIDI-based musical synthesis and similar artificial and random methods of generating music using automata of various kinds.

As well serving purely aesthetic (or recreational) demands, much research has been carried out into pragmatic uses. This includes the representation of stock market data according to pitch for the blind, enhancing usability of information used in applications for use with small screen devices such as PDA's and mobile phones, and sonification of remote vehicle operation, such as those used by bomb disposal teams [6]. However, an application area that has not received much attention is the use of sonic event authentication approaches.

The term authentication refers to the process of ensuring the identity a user asserts is genuine and valid in order for them to access restricted information or services. Well established methods include, for instance, alphanumeric passwords where a user must accurately recall a selected word, phrase or string of characters and/or digits to gain access, or biometrics where only the individual who possesses a unique biological or behavioural characteristic can gain access.

Associated with both schemes are a number of significant drawbacks. Alphanumeric passwords are subject to a trade-off between security and memorability, resulting in users selecting weak, easily subjugated passwords that are easily remembered [9] or selecting strong passwords that are so arbitrary and lengthy in nature that they are extremely difficult to remember [10] especially when used infrequently [13]. The alphabet size is restricted by the number of printable characters on the entry device (e.g. keypad or keyboard). Thus, password sequences are prone to what are colloquially termed "brute force" attacks, where every possible combination of characters are tried until a valid login sequence is found, or "dictionary attacks" where words in a dictionary, along with commonly used variations are used in an attempt to discover login accounts that can be compromised [9]. Biometrics (and in fact all "classical" authentication schemes) are too prone to weaknesses resulting from the nature of their underlying alphabets, (further brief discussion of this follows in section two). It is our aim to show that computer generated music can be used to address some of the problems associated with authentication in general. Thus, successfully utilizing the inherent aesthetic properties of music to fit the intended purpose in a way that classical authentication schemes lacking in these qualities are unable to equal.

The remainder of this paper is organized as follows. In the next section we motivate the use of alternative authentication systems by a short analysis of the two main systems in use today. Namely biometric and password-based systems. In section three we critically discuss image-based password systems. These are well researched and conceptually the closest to our envisaged sound-based systems. The analysis made in the sections two and three is then used in section four to derive the demands on a sound-based authentication system. Our example system is then described in section five. In the following section six, before the conclusion, we discuss how the demands created in section four are met by our system..

2. Alphanumeric passwords and biometrics

A good authentication scheme must be scalable and flexible enough to cope with changing demands in use,

secure, not prone to problems related to "trust", maximise on cognitive strengths and support weaknesses especially in terms of memorability. In addition such schemes should ideally match cultural and environmental needs and expectations and supply a favourable cost-benefit return particularly in terms of time to authenticate and initial cost of implementation.

Alphanumeric passwords are also prone to observation attacks which can come in many guises, from simple "shoulder surfing" where an inauthentic user looks on as the password is entered noting the keys pressed [4], [12], to more complex methods such as training a computer to deduce which keys are pressed during a login session by "listening" to the minute differences in sound that emanate from the keys as they are pressed [1]. Alphanumeric passwords are easily and frequently communicated, making trust a central issue: if a user's trust can be gained, then so it results can their password credentials. These problems are further exacerbated as authentication is becoming so commonplace in every day life that users, faced with multiple services to access tend to reuse passwords between systems in order to cope with the sheer number that they are required to recall. It is therefore often the case that if one password is found, many login accounts can be compromised [8].

Biometrics schemes solve several of these problems. The data they use does not require recollection by the user during authentication. They are not prone to restrictions in underlying alphabet sizes and no keypads or keyboards are used to enter them. They are however, still prone to observation attacks. Biometrics come in two forms, what a person is (physiological) and what a person does (behavioural). Physiological biometrics have to be external to the body in order for them to be as non-intrusive as possible, however this means that they can be observed either directly, for example capturing the image of an iris on video or indirectly, for example "lifting" a fingerprint from a surface that it has been in contact with. This information can then be used to recreate the biometric in order to trick the authentication system [17]. Behavioural biometrics, such as the way a person types can change. This can happen for a variety of reasons, for example with stress, illness or age, which can result in them being unreliable as a means for authentication [3].

3. Media Authentication Systems

Recognition-based authentication schemes are a relatively new concept resulting from further consideration of the inherent drawbacks of existing security technologies. The media to be recognised is any that can be perceived and differentiated by the inexperienced user and that is suitable for intended context of application (for example telephony or over the internet). Images are one type of media that have attracted much interest as the basis for authentication.

Image-based passwords can be thought of in terms of four distinct categories: visual, graphical, visuo-spatial manipulation and visuo-spatial selection. Research suggests that systems of the visual category where a user selects a series of icons, photographic or art images to produce a password value [10] has the potential to meet these requirements [7]. As image-based passwords have been extensively researched, in sharp contrast to sound-based password systems, we give a brief overview on these systems in order to identify the suitable demands on sonic systems.

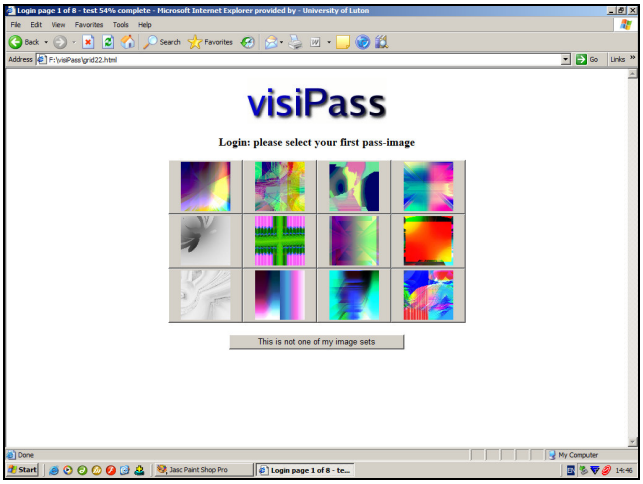

Figure 1. An example of a visual image-based password scheme based on random art

There are a virtually infinite number of images that can be used as underlying alphabets for visual image-based passwords. They are easily based upon computer generated abstract art [2] that though was not developed specifically for this purpose is well suited for transference across networks as images of this type are rendered from a seed value and only this seed needs to be transmitted [4]. Images of objects with pre-existing concrete associations are also difficult to communicate in a format easy for an inauthentic user (trusted or otherwise) to decode into a valid login sequence [4], [7].

Image-based passwords do however have inherent drawbacks of their own in that they cannot be used by those who are blind or partially sighted [6] and are difficult (though not impossible) to be used by those 8% of males and 0.5% of females for example in the UK who have defective colour vision [16]. Image-based passwords are restricted in that they cannot be used in situations where it is not possible to use a screen, such as when authentication is required over the telephone [6]. As with any recognition-based scheme, the explicit nature of the password alphabet means that additional guarding

mechanisms need to be included in order to prohibit observation attacks such as shoulder surfing.

4. Sound-based alphabets

In addition to the overall fundamental requirements discussed so far for authentication systems, sounds used within the system that we propose (termed sonic events) must be distinguishable between, and recognisable by the musically untrained ear. The generated sonic events must be virtually infinite in number in order to guard against brute force attacks. This means they must be generated from a seed (as is the case with abstract art in image-based systems) in order to make fast transmission and storage feasible. Sonic events must also be within the audible frequency range of around 20 Hz to 20 kHz and below the 130 decibel pain threshold. We note however that these values vary between users.

Sounds are differentiated between by pitch, loudness and timbre. Timbre is determined by the acoustic content of a sound (the number and relative intensity of the upper and lower harmonics present in the sound) and the dynamic characteristics of the sound such as vibrato, tremolo and the attack-decay envelope of the sound. In the context of authentication the usable range is much lower. Memorability of sound sequences is a factor of central importance. The scheme that we propose utilises recognition - the sonic event being recognised acts as an external cue. It takes around sixty milliseconds to recognise the timbre of a tone, therefore each sound used in the system must last for at least this length of time. Low harmonics take around a 10 decibel change for us to perceive a change in timbre and around four decibels in mid or high harmonics. Therefore it is also essential that variance between sounds in a sequence must comfort to these values.

Distinctive or unique cues naturally lead to distinct memory traces that are more readily retrievable from long term memory stores during recognition related tasks [5]. This means that by following the above guidelines for sound differentiation sound-based passwords should also remain recognisable even after long periods of time. The system must be implementable in terms of resources such as hardware requirements (we have aimed for a solution that is non hardware specific as there are a number of areas in which sound-based authentication could be used). The system itself should be modular and extensible to allow for integration into existing legacy or newly developed systems. In summary, sonic events for authentication must be:

- Distinguishable by the musically untrained ear.
- Recognisable through physical characteristics.
- Randomly generated from seed values.
- Virtually infinite in number.
- Implementable..

5. Production of sonic events

5.1 Overview

We chose to base our composition on piecewise continuous “well behaved” mathematical functions. See Figure 2 for examples of such functions. Technically these are generated using random cubic polynomials with certain properties. The coefficients of these polynomials are chosen randomly within a certain range that forces a mostly smooth behaviour. Any excessive function values are replaced by a constant (fixed, but randomly determined) term. We note that the fine-tuning of the parameters that determine these polynomials is part of an experimental creative process, targeted to meet the requirements of the sonic events. We wish to point out that "random" here means de facto "pseudo random", i.e. the generated "random values" can be exactly reproduced by the provision of a seed. Let us also emphasise further, that our implementation is only one example from other (possibly similar or different) strategies to meet the demands of sonic authentication systems. The following detailed discussion may encourage the interested reader to find alternative ways to achieve the same goal, i.e. meeting the demands posed by authentication systems.

The compositional process that generates the sonic event consists of a sequence of pitch classes derived from the harmonic chromatic scale within a relatively narrow absolute pitch range. These pitch classes are assigned to a number of voices, that are each determined by three of the mathematical functions described above. This "orchestration" is then used to form a short composition comprising a sequence of single pitches produced by a random choice of one of the voices. Because the duration is also randomly determined, some sonic events are perceived almost as simultaneities (i.e. the ear does not easily detect each and every pitch class/voice so the pitches appear to merge together), whilst others contain few(er) voices. Of course, typically one sonic event contains both perceived simultaneities as well as distinct pitch class events. In the following we explain the compositional process in more detail.

5.2 Voices

A voice is fully described by three functions, a set of instruments and the pitch range. The three functions map time of occurrence to the values pitch, velocity and instrument selection.

The pitch function defines a discrete pitch chosen within the defined absolute pitch range. Although the values of the functions are continuous we prefer to round these values to the nearest semitone of the harmonic chromatic scale. As we will later see, this meets the demands for ease of implementation (implementability) and also (possibly) human memorability too.

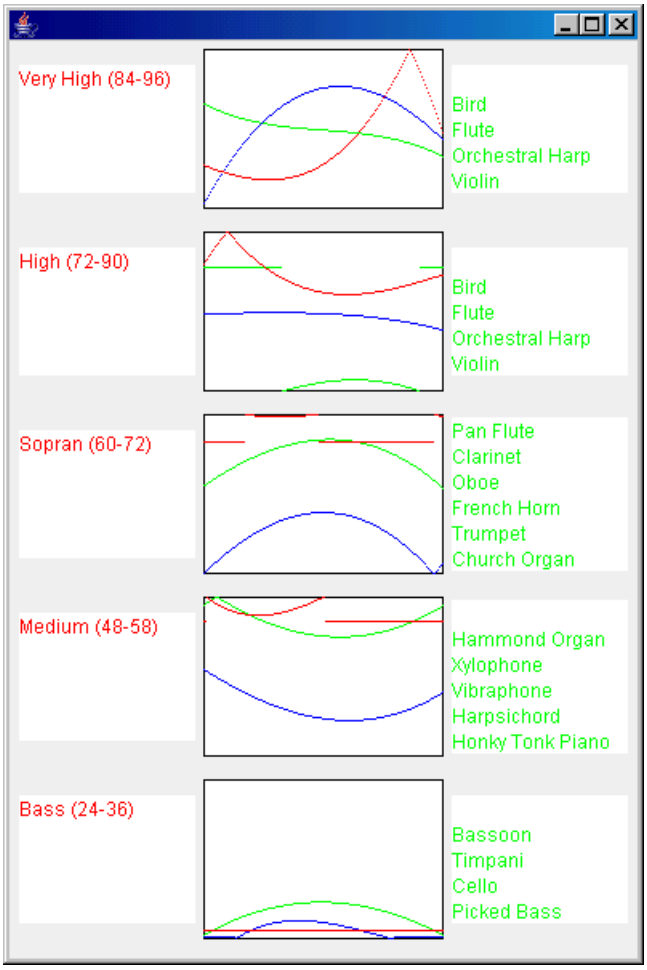

Figure 2. Visualization of Sonic Events

The velocity function defines the velocity as used in the MIDI standard and implemented in Java [14]. Usually the velocity has a different meaning for different instruments. However, as a rule of thumb, a higher velocity creates a "crisper" entrance, giving this instrument slightly more emphasis within the "instrumental mix". A lower velocity creates a less definite entry point.

The last function (instrument selection) defines the choice of the instrument that is used to produce each discrete pitch event within the sonic event. Each voice comes with a collection of predefined instruments. At any given point in time only one instrument per voice is active**.**

5.3 The voice assignment

The full set of musical material available for the composition in our chosen example instrumentation comprises five voices, each acting within their own well defined dedicated absolute pitch ranges. In the following tabulation we use the MIDI enumeration of pitches where 60 = "middle C" and increments or decrements by one denote semitone steps. The instruments are a selection from the default set of instruments provided by the Beatnik Java Soundbank that is shipped as default application with the actual Java distribution.

- Bass Voice. Pitch range: 24-36;
- Medium Voice. Pitch range: 48-58;
- Sopran Voice. Pitch range: 60-72;
- High Voice. Pitch range: 72-90;
- Very High Voice. Pitch range: 84-96;

See Figure 2 for the set of instruments that is assigned to each of the voices.

For each voice our instrument selection (hence instrumental mix) was made so as to increase the ability to clearly distinguish different compositions from one another using varied timbres, thought most likely to be recognized by most non-musicians (brass, strings, etc.). Also, for oral clarity, the lower (Bass) instruments each act within disjunctive pitch ranges. The use of High Voice and Very High Voice with the same set of instruments was motivated by the need to exploit the relatively good perception of the high frequency human hearing area, whilst simultaneously avoiding too many instrumental "flavors". In a sense, our inspiration for this approach can be seen as similar to that of "klangfarbenmelodie", that is to say, creating meaning and memorability to sequences of pitches using instrumental colour as a strong cue, not merely the intervallic structure of the sequence of pitches (which in our case are not merely treated equally as in Schoenberg, but are actually selected according to random mathematical functions without the aid of a human compositor).

5.4 The composition

Each fully populated sonic event is generated by a sequence of up to 120 discrete pitches that are generated at intervals of 5-10 ms. Thus, the duration of each full sonic event is on average 1.2s. In each step of the generation, one of the five voices is randomly selected so as to produce a sound according to the defining functions and parameters of the voice. For clarity of human perception we decided that a sound will only be generated by a voice if there was a change of pitch or instrument. This de facto generates the pitches with different lengths but avoids mere repetition.

Each individual "composition" (sonic event) can be visualized by the diagram shown in Figure 2. Time is - as in any musical score - aligned horizontally. The red line defines the pitch, the blue line velocity and the green line the function that defines the performing instrument.

6. Sonic events for authentication

In section four we developed the key properties for sonic events being used in authentication systems. These are in particular: distinguishable, recognisable, randomly generated, virtually infinite, implementable. In the

following we show how each of these demands are met by our composition strategy.

6.1 Distinguishable by a lay user

As the parameters (i.e. the generating functions that define the composition) are chosen by a random number generator it is nearly impossible to mathematically prove that different seeds in fact generate sonic events that are perceived to be "different" by the lay (non-musically trained) listener. However a number of design decisions have been made so as to maximize distinguishability:

- The pitch ranges of the lower voices are disjunctive, hence ensuring that the instruments are as clearly recognizable as possible by the non-expert listener.
- The sets of instruments for the lower voices are disjoint. An instrument that serves, for instance, the medium voice (e.g. the Church Organ) does not appear as an instrument of one of the other voices.
- The composition can be visualized via mathematical functions. Hence we can at least ensure a certain timbral and pitch class "richness" and heterogeneity of generated sonic events by experimenting with a variety of generating functions. The examples in Figure 2 show that most of the functions, as though they are all suitable, behave completely differently.

In summary, there is within and between each sonic event considerable oral variety created by the random selection of distinctive timbres, each with a random entry (hence exit point) and each operating within their assigned pitch ranges with considerable variety (not only pitch but also number of pitches per voice).

6.2 Recognisable through physical characteristics

The work presented in this paper is meant to be in progress, and extensive human cognitive studies ensuring the practicality of the sonic events as passwords or more precisely, "passphrases" are still to be completed. However, certain measures have already been taken to enhance their potential heterogeneity without compromising memorability.

Although, for usability purposes, the sonic events need to be short, they are long enough so as to leave the cognitive impression of a possibly truncated melody/orchestral fragment. The deliberate choice of familiar instruments (some orchestral, some not) operating within the familiar chromatic harmonic scale was designed to aid memorability and familiarity with other musical contexts. At least in Western cultural terms using "high art" timbral and pitch elements, memorability is potentially enhanced. It should be however noted, that the implementation of sonic events in other cultural contexts (non-Western, popular street cultural forms more familiar to younger urban audiences and multi-cultural audiences) may need to follow therefore a different strategy.

6.3 Random generation from seed values

All random numbers used in the production process of the parameters that determine the voices and the composition are produced by the standard Java pseudo random number generator. [15]. Hence although the values appear and behave like randomly chosen numbers, their behavior is determined by the seed.

6.4 Virtual Infinity of Password Space

If we focus on instrumentation and pitch and ignore velocity for a moment we observe that each voice has at least four instruments and a pitch range of at least 10 semitones leading to more then 200 discrete sounds (pitch class/timbre pairs). Each composition consists of a sequence of 120 of these atomic sonic events, leading to a potential total population of more then 20,000 different compositions.

While there is still a "philosophical" gap between this large potential corpus of musical material and infinity in the strict mathematical sense, this large population of sonic events is sufficient for all practical purposes, in particular when compared with the standard alphabet set of traditional password systems derived from the ninety-five printable keys on an ASCII keyboard.

We wish also to point out, that it would be a relatively easy task to enlarge this potential musical input set even more, for instance by changing the set of instruments used, adding MIDI controllers as Pedal, Tremolo, etc. as part of the compositional process. However, for demonstration purposes we decided to keep the system as described here relatively simple and straightforward so as to provide the basis for human perception trials of the application of the sonic event to its intended purpose.

6.5 Implementability

The application areas of the sonic events are authentication systems. Hence they do not exist as isolated “music producing hard- or software”, but are likely to be embedded as part of a wider system that meets industrial standards. Hence our decision was to use techniques that are generic to any of today's synthesizers.

In fact, our system has been implemented in 100% pure Java (version 1.5) and the sound is generated by the Java default synthesizer using the Beatnik Soundbank that is distributed as part of the Java runtime environment. The effect of this is that it is available on any computer that is able to run Java programs.

In addition, the modular, object oriented code design allows the direct interaction with any other Java code, or, using standard distributed techniques, implementation of the sonic events in form of a software service interacting with other programs.

7. Conclusion

Almost from its emergence as a viable research and pragmatic tool (from the late 1940's onwards) the modern computer has been harnessed by humans so as to generate,

analyse, synthesise, hence deepen our understanding of modern as well as more ancient musical forms. One of our aims is to show that even the most pragmatic applications required in the modern world (i.e. the need for secure access to workstations and computer networks in offices) can be successfully addressed with computer generated sonic events that contain just enough expressivity and auditory familiarity to be described as primitive compositions.

Indeed, by harnessing the familiar dimensions of "high-art" Western music (familiar harmonic scales and instrumentation) but within a different pattern of constraints and without the human mind as an artistic creator / editor) it is possible to meet the pragmatic needs of authentication (memorability, etc.) without compromising some degree of aesthetic beauty or quality.

It is by engendering the events with precisely these aesthetic qualities that we may hope lead to systems that will better fit their intended purpose, indeed far better than the existing alternatives, such as classical alphanumeric passwords that lack these aesthetic dimensions. This is not to say we are claiming to invent a new type of musical composition or to rival the acknowledged experts in computer music or MIDI creativity. Rather, that in marrying some elements of "high-art" with pragmatism we were able to create a useful application that is pleasing to the ear and contains some musical (as compared to purely sonic or random noise) elements. To achieve an optimal balance in this particular context of application may require further experimentation and re-engineering. It is our purpose in writing this paper to instantiate further activity in this area and also encourage others to consider using sounds in new areas of application such as in this case, a tool for authentication.

8. References

[1] Asonov, D., Agrawal, R. Keyboard Acoustic Emanations. Proceedings of the 2004 IEEE Symposium on Security and Privacy (S&P '04), Oakland, CA (2004)

[2] Bauer, A. Andrej Bauer's random art website: http://www.random-art.org (1998)

[3] Boertien, N., Middelkoop, E. Authentication in mobile applications, CMG Telematica Instituut, available from https://doc.telin.nl/dscgi/ds.py/View/Collection-4122 (2002)

[4] Dhamija, R. and Perrig, A. Déjà Vu: A user study using images for authentication. In Proceedings of the 9th USENIX security symposium, Denver, Colorado, Aug. 14-17 (2002).

[5] Eysenck, M.W., and Eysenck, M.C. Effects of processing depth, distinctiveness, and word frequency on attention. British journal of psychology, 71, 263-274. (1980)

[6] Franklin, K., and Roberts, J. A Path Based Model for Sonification. In: Ebad Banissi et al, (Ed), Information Visualization, pages 865-870. IEEE Computer Society, July 2004.

[7] Gibson, M., Conrad, M., Maple, C. A Methodology for Evaluating the End-user Acceptance of Image-Based Passwords, submitted for publication, available as preprint from http://perisic.com/preprints (2006).

[8] Kent, S., and Millet, L. (Ed). Who goes there? - Authentication through the lens of privacy, Washington, USA., The National Academies Press (2003). Available at: http://books.nap.edu/html/whogoes

[9] Klein, D. Foiling the Cracker: A Survey of, and Improvements to, Password Security, Proceedings of the 2nd USENIX Unix Security Workshop, Oakland, CA, August, 1990, pp.5-14 (1990).

[10] Jansen, W. Authenticating Users on Handheld Devices, National Institute of Standards & Technology, Gaithersburg, Maryland, USA. [Online] Available from: http://csrc.nist.gov/mobilesecurity/Publications/PP-AuthenticatingUsersOnPDAs.pdf (2003)

[11] Maple, C., Conrad, M., French, T. Towards More Trustworthy B2C E-commerce Using a Hybrid Authentication Scheme. In Procs. IADIS International eSociety Conference Vol. 1, p 134-141, Avila, Portugal (2004)

[12] Mitnick, K. and Simon, W. The Art of deception: Controlling the Human Element of Security, New York, NY, John Wiley & Sons, Inc. (2002)

[13] Sapolsky, R. Stressed out memories, Scientific American Mind, Vol 14,(5) (2004)

[14] Sun Microsystems, Documentation of the interface MidiChannel in Java 2 Platform Standard Ed. 5.0. Available at http://java.sun.com/j2se/1.5.0/docs/api/javax/sound/midi/MidiChannel.html (2005)

[15] Sun Microsystems, Documentation of the class Random in Java 2 Platform Standard Ed. 5.0. Available at http://java.sun.com/j2se/1.5.0/docs/api/java/util/Random.html (2005)

[16] Voke, J. The industrial consequences of deficiencies of colour vision, PhD thesis, The City University, London (1976)

[17] Weinshall, D., and Kirkpatrick, S. Passwords you'll never forget but can't recall, CHI 2004, April 24-29 (2004)