MEASURING THE HUMAN VOICE: ANALYSING PITCH, TIMING, LOUDNESS AND VOICE QUALITY IN MOTHER/INFANT COMMUNICATION

Stephen N. Malloch (1), David B. Sharp (2), A. Murray Campbell (2),
D. Murray Campbell (2), Colwyn Trevarthen (1)

University of Edinburgh:
(1) Department of Psychology, 7 George Square, Edinburgh EH8 9JZ, Scotland.
(2) Department of Physics, JCMB, Kings Buildings, Mayfield Road, Edinburgh EH9 3JZ, Scotland

e-mail: stephen.malloch@ed.ac.uk

1. INTRODUCTION

A mother and her young baby are playfully interacting. We hear the mother speak in short bursts, talking in a sing-song manner, and the baby occasionally 'answers back'. It appears clear that communication is taking place, but communication based in what ? The baby cannot understand the meaning of the words the mother is using, and the baby answers in often 'gliding-type' sounds. The communication must be 'held' by means other than lexical meaning, grammar and syntax.

By combining the disciplines of psychology, psychoacoustics and music, we attempt to shed light on what takes place when a mother and her infant vocalise together in this manifestly communicative way.

2. BACKGROUND

When mothers speak to infants, their voicing appears to have strong metrical and melodic characteristics, and this song-like activity seems to be attended to and responded to with much pleasure by infants. Indeed, infants often stimulate an affectionate adult, male or female, to extended poetic or musical speech, as well as other types of music-like sound making. This distinctive style of adult speech is called 'motherese,' 'parentese,' or 'infant directed speech' - it varies with the age and state, and motives and emotions of the infant partner.

Research on mother-infant communication, in which micro-analyses have been made of the behaviours involved, and experimental studies done of infants' reactions to different elements of human expression, has revealed that infants possess complex endowments for perceiving and stimulating maternal communicative signals. Infants can discriminate timing patterns, pitch, loudness, harmonic interval and voice quality [1]. These abilities emerge prenatally. It has been found that an infant learns its mother's voice from before birth, and can recognise melodies or poetic verses that were presented for it to hear prenatally. Reactions of new-borns to the human voice and their imitations of facial expressions, vocalisations and hand movements, prove that this awareness of human signals, while slow and rudimentary, is both comprehensive and coherent at birth [2].

Complementary research on mothers' expressions when they are addressing their infants proves that the innate system involves special abilities on both sides. Mothers' speech to infants has unconscious or intuitive form that has many of the same characteristics in all languages. The tone of a mother's voice (the 'voice quality'), and its rhythms and melody, are all regulated in predictable ways. These features match the demonstrated preferences that young infants seek in a human partner. Typically, mothers repeat short, evenly spaced words with simple, sing-song intonations in a resonant yet relaxed and 'breathy' moderately high-pitched voice [3]. Baby and mother listen to one another's sounds, creating co-operative patterns of vocalisations

3. METHODS OF MEASUREMENT

In order to understand more fully what occurs when a mother and her infant vocalise with one another, and to elucidate its probable function in development, we need ways of measuring the vocal interaction with adequate precision.

Six methods intended to help us perceive this interaction in detail are discussed in this study - spectrographic analysis, pitch plotting, loudness measurement, sharpness measurement, roughness measurement, and tristimulus values. Our aim is to find a set of complementary windows into mother/infant vocal communication, and the mutual regulation of the complementary parts of mother and infant vocalisations.

3.1 Spectrographic Analysis

A spectrograph allows a view of vocalisations through time. Figure 1 is a spectrographic analysis of Laura, a 6 week old infant, and her mother vocalising together. This is a 10 second excerpt from an analysed interaction lasting 30 seconds. The transform size of the Fast Fourier Transform was 4096, the overlap was 2048, and a Hamming window was used. Intensity is shown by a grey scale, the calibration of which is shown in the top left of the diagram.

The bracketed letters correspond to utterances by the mother: (a) "come on"; (b) "again"; (c) "come on then"; (d) "that's clever"; (e) "oh yes"; (f) "is that right". The boxed section shows the infant's vocalisation. The interpretation of this information and the significance of the vertical 'bar-lines' will be discussed in section 4.

3.2 Pitch Plots

Spectrographs are not good at revealing high resolution fundamental frequency information, especially at lower frequencies. In order to investigate how the mother and baby explore frequency space during their interactions, a method is needed for graphing the pitch of the vocalisations.

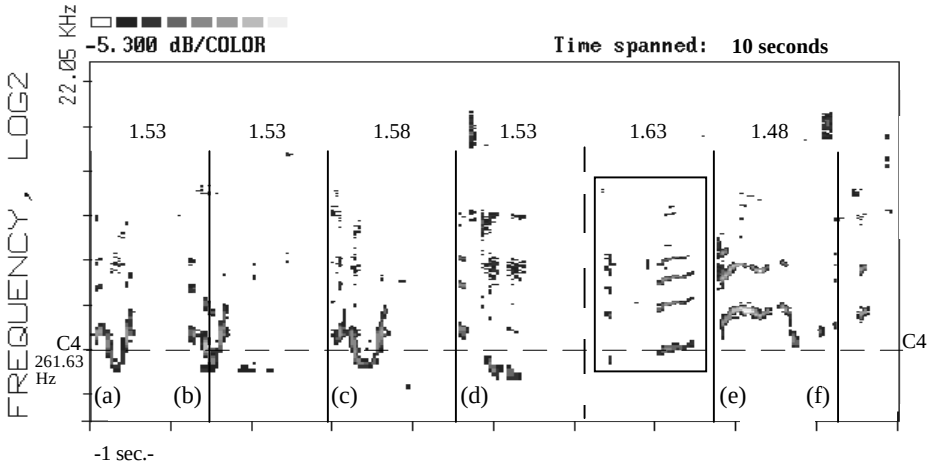
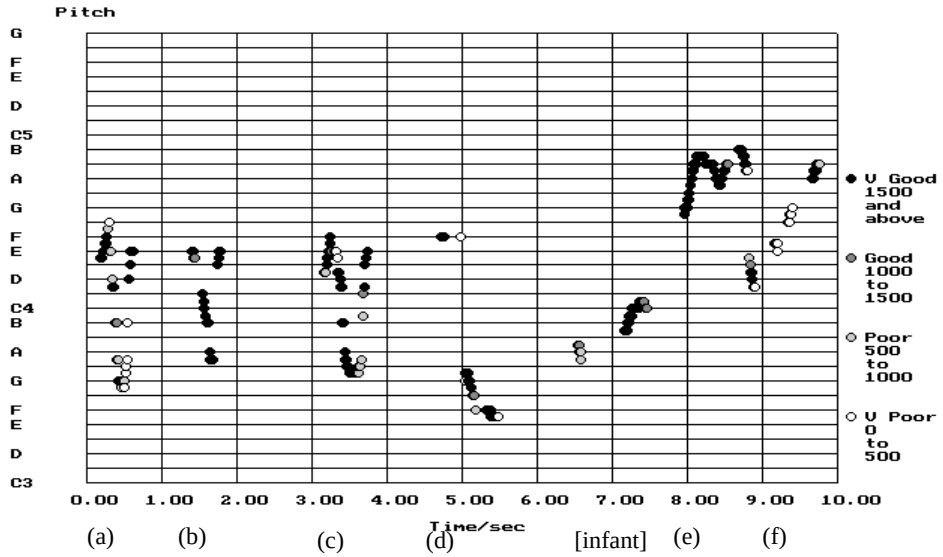


Figure 1

In Brown [4], the calculation of a constant Q spectral transform is described that produces a constant pattern in the log frequency domain for sounds with harmonic frequency components. In Brown [5], a method is described whereby this log frequency pattern is correlated with an ideal harmonic pattern, and thus the fundamental frequency can be determined. Brown's methods have been implemented by us in software, enabling the pitch of the vocalisations to be plotted.

Figure 2 shows a pitch plot of the data represented in figure 1. A data point is calculated every 0.01 seconds, and the frequency resolution is 1/4 of a tone. Pitch is represented on the vertical axis (in letter names) and time is represented along the horizontal axis. As in figure 1, bracketed letters, which refer to the mother's words, are placed below the graph.

The audio data is split-up into time-windows (in this instance, 0.39 seconds long), and a constant Q transform performed on each. The software searches each time-window for the strongest correlation with a 'template' composed of a set of harmonic relationships. This template is defined as consisting of six harmonics progressing in amplitude from 1 (1st harmonic) to 2/3 (6th harmonic) in equal steps. The strongest correlation within each window is defined as the fundamental frequency. Because not all sounds will be pitched, a 'correlation floor' can be defined so as to filter out very weak correlations that will have no foundation in perception. Figure 2 also shows the strength of correlation for each pitched event - the key is shown on the right of the graph.



(a) (b) (c) (d) [infant] (e) (f)

Figure 2

3.3 Loudness Plots

We calculate loudness, measured in sones, either by Stevens' Mark VII Procedure [6] (used in Pollard and Janson's Tristimulus calculations: see section 3.6) or by the method proposed by Zwicker [7] (used in Aures' Sharpness calculations: see section 3.4). Because a recording has no intrinsic level of loudness, the calculation procedure requires a value for loudness to be entered which represents the listener's notion of the loudness (in sones) of the loudest moment in the section under analysis.

3.4 Sharpness Plots

Sharpness, measured in acums, is calculated according to the formula proposed by Aures [8]. Sharpness is related to the position of the loudness centroid in a sound's spectrum.

3.5 Roughness Plots

Roughness is caused by beating between partials. The model of roughness measurement we use is that proposed by Hutchinson and Knopoff [9]. While Hutchinson and Knopoff's model is fairly rudimentary, it is our opinion that no substantially better model yet exists.

3.6 Tristimulus Values

The Tristimulus Method of timbre measurement was developed by Pollard and Janson [10]. It compares the relative loudness of three spectral areas of a harmonic sound - the fundamental, harmonics 2-4, and harmonics 5-n. Usually, the results are plotted within a triangular space (which omits a time axis), though we have found that the results can also be successfully represented as a tripartite division of vertical bars placed along a time axis.

4. INTERPRETATION

4.1 Spectrographic Analysis (figure 1)

By viewing vocalisations placed in time, rhythmic patterns become easier to identify. What was at first an intuitive judgement that a regular beat could be discerned during the course of listening to the 30 second interaction, is confirmed by the identification of 'bar-lines' (the number above each bar is its length in seconds). Identifying the bar-lines by the letter closest to them, bar-lines ***a***, ***c***, ***d*** and ***e*** occur at the onset of an utterance; bar-line ***b*** occurs at the lowest part of a pitch bend, and bar-line ***f*** occurs at an emphasised word ("is *that* right"). The dashed-line bar-line occurs between events - the time-interval from bar-line ***d*** to this dashed bar-line represents the most common bar-length in this 30 second interaction. Here, the mother allows space for the infant to reply. These bars show a remarkable regularity, and the whole interaction shows that the mother and infant enter into co-ordinated communication.

The periodicity of this bar-structure is further demonstrated from the results of performing a 4k FFT on the loudness data of this 30 second excerpt. The results show that the highest spike lies at around 0.3 s^{-1}, with a lesser spike around 0.6 s^{-1}. This demonstrates periodicities of around 3.3 and 1.6 seconds. Given the non-exact nature of the bar structure, and that the bar-lines sometimes lie *between* events, this result further strengthens our initial belief of a bar-structure of approximately 1.53 seconds - evidence of a regularity that allows the mother and infant to negotiate their turn-taking.

4.2 Pitch Plots (figure 2)
Pitch plots allow us to observe how the pitches of the mother's and infant's vocalisations progress during the course of an interaction. The U-shaped pitch curves of vocalisations (a), (b) and (c) are characteristic of a vocalisation that invites activity. It appears that mother and infant explore pitch-space in a methodical manner. The examination of pitch plots over a span of 5 minutes reveals regular 'waves' of pitch movement by the mother, the infant following this trend in its own vocal pitches. We have also found evidence of very precise pitch (and rhythm) matching by infants to their mother's vocalisations (both between and on top of maternal vocalisations). These infant vocalisations are often 'musically logical' - particularly during songs sung by the mother. Pitch (and loudness) movements by the mother and infant will have important emotional and motivational roles within their co-operative interaction.

4.3 Voice Quality
Measures of Sharpness, Roughness and Tristimulus values are all measures of voice quality, or vocal timbre. By analysing these measures, it becomes possible to observe in detail how the voice quality of the mother changes during different types of behaviour, and at different ages of the infant.

The same mother represented in the previous figures uses the phrase "come on then" seven times during a recorded session of her interaction with Laura, aged 4 weeks. Figure 3 shows the difference in tristimulus values for six occurrences of this phrase (utterance number 5 contains no identifiable fundamental, so cannot be represented by tristimulus values).

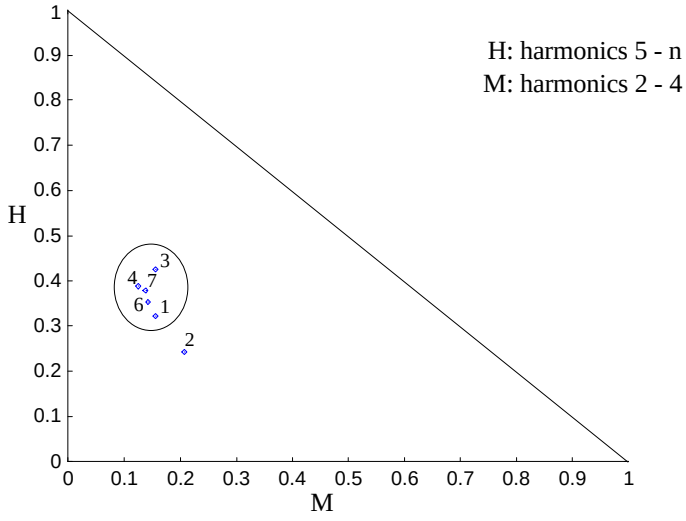


Figure 3

The points on this graph represent averaged tristimulus values for each utterance of "come on then." The horizontal axis measures the relative loudness of the middle harmonics (2-4), and the vertical axis measures the relative loudness of the upper harmonics (5-n). Through independent subjective voice quality assessment (by a person trained in Laver's voice quality assessment procedure [11]), it appears that tristimulus values correlate well with a measure of laryngeal tension. In figure 3, the points that are circled represent utterances that are, overall, lax and contain whisper (the point labelled 3 being the quietest). Point 2, on the other hand, is tense, and contains no whisper (and is the loudest).

We can obtain a fuller understanding of the quality of each utterance by combining psychoacoustic measures. Figure 4 contains data on pitch, loudness, sharpness and roughness for the seven instances of "come on then" discussed above.

Utterance number 2, as well as lying in a different area on the tristimulus graph from the other utterances, relatively exhibits the most consistently high pitch, is the loudest, and shows low measures of sharpness and roughness. Thus, an 'auditory profile' can be built-up, and the interpersonal function of this profile can then be judged from an examination of its communicative context.

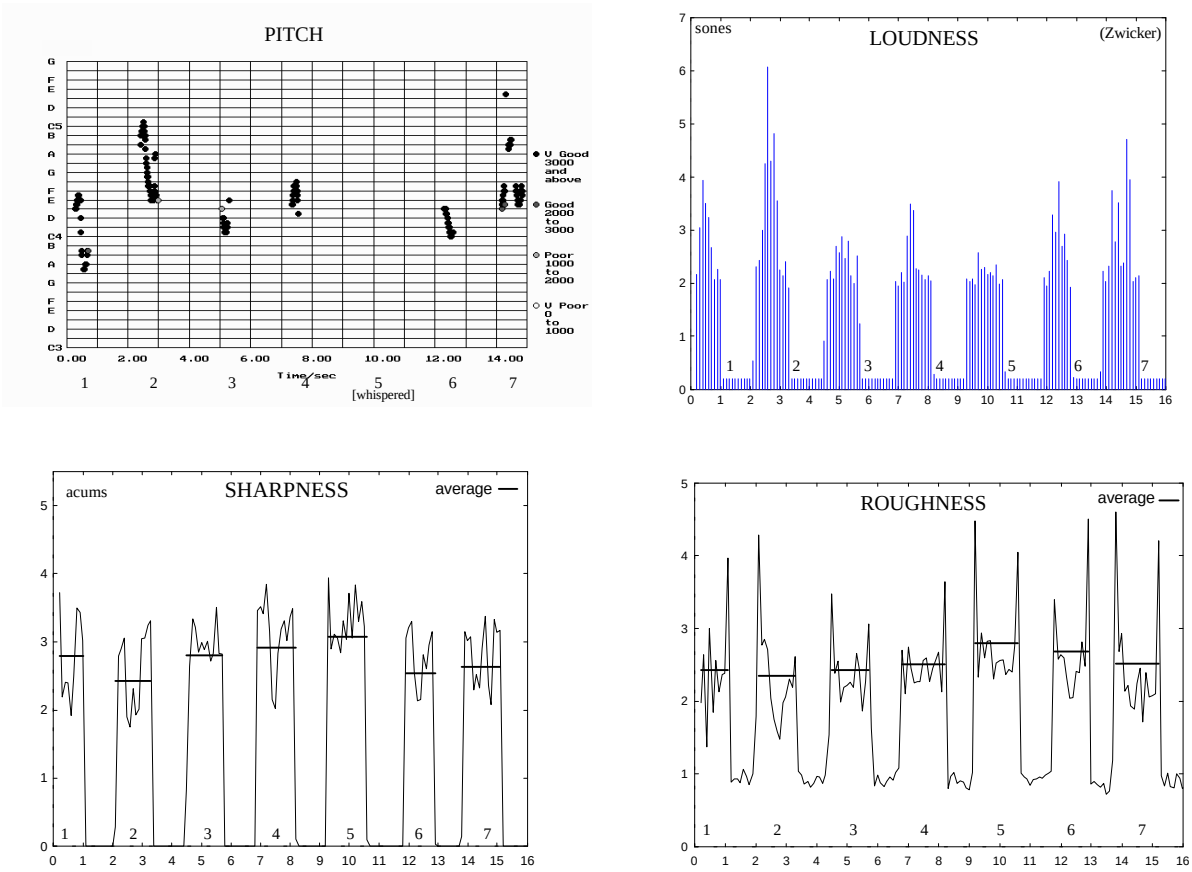


Figure 4

5. CONCLUSIONS

We have outlined six acoustic measures which we believe reveal pertinent information about the communicative nature of vocalisations between mothers and infants. Further work will be carried out examining the intersection of Laver's subjective voice quality assessment techniques and the acoustic measures presented here, and we will examine in detail the communicative contexts of the analysed utterances. The sound-world of the mother and infant tells us much about the communicative and emotional potential of human vocalisation.

6. REFERENCES

[1] Trehub, S. E.,Trainor, L. J. and Unyk, A. M. (1993) "Music and speech processing in the first year of life". *Advances in Child Development and Behaviour* 24: 1-35.

[2] Trevarthen, C., Kokkinaki, T., and Fiamenghi, G. A. Jr. (1997) "What infants' imitations communicate: with mothers, with fathers and with peers." In J. Nadel and G. Butterworth (Eds.), *Imitation in infancy*. Cambridge: Cambridge University Press. (in press).

[3] Stern, D.N. et al (1982) "Intonation contours as signals in maternal speech to prelinguistic infants." *Developmental Psychology* 18: 727-735.

[4] Brown, J. C. (1991) "Calculation of a constant Q spectral transform." *J.Acoust.Soc.Am*.89:425-434.

[5] Brown, J.C. (1992) "Musical fundamental frequency tracking using pattern recognition method." *J. Acoust. Soc. Am* 92: 1394-1402.

[6] Stevens, S.S.: (1972). "Perceived Level of Noise by Mark VII and Decibels (*E*)." *J.Acoust Soc Am* 51/2: 575-601.
[7] Paulus, E., & Zwicker, E. (1972) "Programme zur automatischen Bestimmung der Lautheit aus Terzpegeln oder Frequenzgruppenpegeln." *Acustica* 27: 253-266.
[8] Aures, W. (1985) "Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale." *Acustica* 59: 130-141.
[9] Hutchinson, W. & Knopoff, L. (1978)"The acoustic component of Western consonance." *Interface* 7:1.
[10] Pollard, H. & Janson, E. V. (1982) "A tristimulus method for the specification of musical timbre." *Acustica* 51: 162 - 171.
[11] Laver,J.(1980) *The Phonetic Description of Voice Quality*. Cambridge, Cambridge University Press.

ACKNOWLEDGEMENTS: Dr. Helen Marwick undertook the subjective voice quality assessment. This research has been funded by a Research Grant from The Leverhulme Foundation.

This paper was presented at *The International Symposium of Musical Acoustics*, Edinburgh, 19-22 August, 1997. It appears in *Proceedings of the Institute of Acoustics*, vol.19, part 5: 495-500.