Proceedings of the

1st International Workshop on

Vocal Interactivity in-and-between Humans, Animals and Robots

VIHAR 2017

Skövde, Sweden, 25-26 August 2017

Published by:
Ricard Marxer

ISBN: 978-2-9562029-0-5

Credits:
Editors: Angela Dassow, Ricard Marxer, Roger K. Moore
Cover photo: Jarke, Skövde from city hall, Recolored by Ricard Marxer, CC BY-SA 4.0,
https://commons.wikimedia.org/wiki/File:Skövde_from_city_hall.jpg
Proceedings assembled by: Ricard Marxer

Workshop took place in Skvöde, Sweden — August 25-26, 2017

Published online at http://vihar-2017.vihar.org/ — September 14, 2017

Phonetic Characteristics of Domestic Cat Vocalisations

Susanne Schötz[1], *Joost van de Weijer*[1], *Robert Eklund*[2]

[1] Lund University, Sweden
[2] Linköping University, Sweden

susanne.schotz@med.lu.se, vdweijer@ling.lu.se, robert.eklund@liu.se

1. Introduction

The cat (*Felis catus,* Linneaus 1758) has lived around or with humans for at least 10,000 years, and is now one of the most popular pets of the world with more than 600 million individuals [1], [2]. Domestic cats have developed a more extensive, variable and complex vocal repertoire than most other members of the Carnivora, which may be explained by their social organisation, their nocturnal activity and the long period of association between mother and young [3]. Still, we know surprisingly little about the phonetic characteristics of these sounds, and about the interaction between cats and humans.

Members of the research project Melody in human–cat communication (Meowsic) investigate the prosodic characteristics of cat vocalisations as well as the communication between human and cat. The first step includes a categorisation of cat vocalisations. In the next step it will be investigated how humans perceive the vocal signals of domestic cats. This paper presents an outline of the project which has only recently started.

1.1. Previous studies

The phonetic characteristics of domestic cat vocalisations were first described by Moelk [4], and since then a number of acoustic characteristics of cat vocalisations have been described [5]–[12]. Based on previous descriptions as well as analysis of new recordings, an attempt was made to develop a comprehensive phonetic typology of domestic cat vocalisations, with phonetic definitions. Table 1 shows the number of vocalisation types and subtypes identified so far.

Table 1: *The most common domestic cat vocalisation types and subtypes identified in this study.*

Vocalisation type	Subtypes
Meow	Mew, Squeak, Moan, Meow, Trill-meow
Purr	-
Trill	Chirrup, Grunt, Trill-meow
Howl	Howl, Howl-growl
Growl	Growl, Howl-growl
Hiss	Hiss, Spit
Snarl	-
Chirp	Chirp, Chatter

Auditory as well as acoustic analyses have been used to identify and describe the different types. The descriptions include phonetic transcriptions, segmental and prosodic features, as well as typical contexts in which the vocalisations are used. These types are now used in the project for annotating and classifying cat vocalisations (see Figure 1 for an example waveform, spectrogram and fundamental frequency (F_0) contour of a vocalisation, and http://meowsic.info for additional video and audio examples.

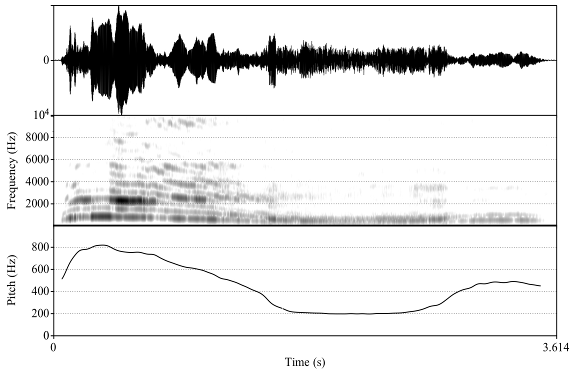

Figure 1: *Waveform (top), spectrogram (mid, bandwidth: 300 Hz) and F_0 contour (bottom) of an example howl-growl.*

2. Vocalisation types

The following list is an overview of the vocalisation types we have identified so far along with their subtypes. Example phonetic transcriptions and typical contexts in which the vocalisation types are used are provided for each type.

1. Sounds produced with the mouth closed
    a. **Purr(ing)**: a low-pitched regular and probably nasalised sound produced during alternating (pulmonic) egressive and ingressive airstream: [↓hːr̃-↑r̃ːh-↓hːr̃-↑r̃ːh…] or ; used when the cat is content, hungry, stressed, in pain, gives birth or is dying; probably signals ”I do not pose a threat” or ”Keep on doing what you are doing”.
    b. **trill (chirr, chirrup, grunt, murmur)**: a short and often soft, sometimes a bit harsh nasalised sound rolled on the tongue, i.e. a voiced trill: [mhr̃ː], [mːr̃ːut], [bʀ̃ːh]; used e.g. during friendly approach and greeting, and during play; grunts (murmurs) are usually more low-pitched, while trills or chirr(ups) are more high-pitched; sometimes cats combine a trill with a meow, producing the more complex vocalisation subtype trill-meow
2. Sounds produced with an opening-closing mouth
    a. **meow (miaow) sounds**: Meows can be assertive, plaintive, friendly, bold, welcoming, attention soliciting, demanding, or complaining, sad or even silent. A meow can be varied almost endlessly, and there are several subtypes, including the following:

i. **mew**: a high-pitched meow with [i], [ɪ] or [e] quality: [mi], [wɪ] or [mɪu]; kittens may use it to solicit attention from their mother, and adult cats may use it when they are sad or in distress or when they signal submissiveness

ii. **squeak**: raspy, nasal, high-pitched and often short mew-like call, sometimes with an [ɛ] vowel quality: [wæ], [mɛ] or [ɛʊ], sometimes not ending with a closing mouth; often used in friendly requests

iii. **moan**: with [o] or [u] vowels: [moau] or [mæu]; often used when sad or demanding

iv. **meow (miaow)**: a combination of vowels resulting in the characteristic [iau] sequence: [miau], [ɛau] or [waʊ]; often used in cat-human communication to solicit food or get past an obstacle (e.g. a closed door or window); adult cats mainly meow to humans, and seldom to other cats, so adult meow could be a post-domestication extension of mewing by kittens

b. **trill-meow (murmur-meow)**: combination of a trill (murmur) and a meow: [mrhɪau], [mhrŋ-au] or [whrrrau]; used in the same contexts as the meow

c. **howl (yowl, moan, anger wail)**: long and often repeated sequences of extended vocalic sounds – often with [ɪ], [ɨ], [j], [ɤ], [aʊ], [ɛɔ], [aw], [ɔɪ], [ɑo] – usually produced by gradually opening the mouth wider and closing it again; used in threatening situations, and often merged or combined with by growls in long sequences with slowly varying F_0 and intensity: [gʀːawɪjɑoʀː]

d. **mating cry (mating call)**: long sequences of meow-like sounds, sometimes similar to the cries of human infants; often used in spring during the mating season: [wa͡ːuw], [ɹːɪːa͡uː], [mhrːwaːoːuːɪː] or [ʀːwːuːaːu]

3. Sounds produced with an open tense mouth are often associated with either offensive or defensive aggression, but also with prey-directed vocalisations

a. **growl (snarl):** long guttural, harsh, regularly and rapidly pulse-modulated, low-pitched sounds produced during a slow steady exhalation, often with the lip curled up and exposed teeth [gʀː], with a vocalic [ɹː] or rhotic [ʌ], occasionally beginning with an [m]; used to signal danger or to warn or scare off an opponent, and often intertwined or merged with howls and hisses

b. **hiss and spit** (the more intense variation): agonistic (aggressive and defensive) sounds produced with the mouth wide open and the teeth exposed, sounding a bit like long exhalations: [hː], [ħː], [çː], [ʃː] or [s̢ː]; often an involuntary reaction to being surprised by an (apparent) enemy; the cat changes position with a startle and breath is being forced rapidly through the slightly open mouth before stopping suddenly; the spit sounds similar to a hiss, but may sometimes begin with a stop – often a t-like sound: [t͡ʃː], [t͡s̢ː], [k͡h̯ː]

c. **snarl (scream, cry, pain shriek)**: loud, harsh and high-pitched vocalic sounds, often with [a], [æ], [aʊ] or [ɛo] vowels: [æːo̰]; often produced just before or during active fighting, or when in pain

d. **chirp and chatter (prey-directed sounds)**: a hunting instinct where cats copy the calls of their prey, e.g. when a bird or insect catches their attention (by making a sound) and the cat becomes riveted to the prey, and starts to chirp, tweet and chatter:

i. **chatter (teeth chattering)**: unvoiced very quick stuttering or clicking sequences of sounds with the jaws juddering, [k̟̚ k̟̚ k̟̚ k̟̚ k̟̚ k̟̚]

ii. **chirp:** voiced short calls said to be mimicking a bird or rodent chirp, sound similar to a high-pitched phone ring, tone often rises near the end, [ʔə] or reiterated [ʔɛʔɛʔɛ...]

iii. **tweet and tweedle:** tweets are soft weak chirps, often without any clear initial [ʔ] and with varying vowel qualities: [wi] or [ɦɛu]; tweedles are prolonged chirps or tweets with some voice modulation, like tremor or quaver: [ʔəɛəɥə]

Previous pilot studies have revealed that experienced human listeners are fairly good at recognizing the vocal signals of domestic cats [13], [14]. In future studies we intend to investigate this further.

3. Acknowledgements

The authors gratefully acknowledge the Marcus and Amalia Wallenberg Foundation and Lund University Humanities Lab.

4. References

[1] C. A. Driscoll, J. Clutton-Brock, A. C. Kitchener, and S. J. O'Brien, 'The Taming of the cat', *Sci. Am.*, vol. 300, pp. 68–75, 2009.

[2] D. C. Turner and P. P. G. Bateson, Eds., *The domestic cat: the biology of its behaviour*. Cambridge University Press, 2000.

[3] J. Bradshaw, R. A. Casey, and S. L. Brown, *The behaviour of the domestic cat*. CABI, 2012.

[4] M. Moelk, 'Vocalizing in the House-Cat; A Phonetic and Functional Study', *Am. J. Psychol.*, vol. 57, pp. 184–205, 1944.

[5] K. A. Brown, J. S. Buchwald, J. R. Johnson, and D. J. Mikolich, 'Vocalization in the cat and kitten', *Dev. Psychobiol.*, vol. 11, pp. 559–570, 1978.

[6] P. E. McKinley, *Cluster Analysis of the Domestic Cat's Vocal Repertoire*. University of Maryland, 1982.

[7] N. Nicastro, 'Perceptual and Acoustic Evidence for Species-Level Differences in Meow Vocalizations by Domestic Cats (Felis catus) and African Wild Cats (Felis silvestris lybica).', *J. Comp. Psychol.*, vol. 118, no. 3, pp. 287–296, 2004.

[8] S. C. Yeon *et al.*, 'Differences between vocalization evoked by social stimuli in feral cats and house cats', *Behav. Processes*, vol. 87, no. 2, pp. 183–189, 2011.

[9] S. Schötz and R. Eklund, 'A comparative acoustic analysis of purring in four cats', in *Proceedings of Fonetik 2011*, Stockholm, 2011, pp. 9–12.

[10] S. Schötz, 'A phonetic pilot study of vocalisations in three cats', in *Proceedings of Fonetik 2012*, University of Gothenburg, 2012, pp. 45–48.

[11] S. Schötz, 'A phonetic pilot study of chirp, chatter, tweet and tweedle in three domestic cats', in *Proceedings of Fonetik 2013*, Linköping University, 2013, pp. 65–68.

[12] S. Schötz, 'Agonistic Vocalisations in Domestic Cats : A Case Study', in *Working Papers*, 2015, vol. 55, pp. 85–90.

[13] S. Schötz and J. van de Weijer, 'A Study of Human Perception of Intonation in Domestic Cat Meows', in *Proceedings of Speech Prosody*, Dublin, 2014.

[14] S. Schötz, 'A pilot study of human perception of emotions from domestic cat vocalisations', in *Proceedings from Fonetik 2014*, Stockholm University, 2014, pp. 95–100.

Proc. 1st Intl. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR), Skvöde, Sweden, 25-26 Aug 2017

Appropriate Voices for Artefacts: Some Key Insights

Roger K. Moore

Speech & Hearing Research Group, Dept. Computer Science, University of Sheffield, UK
r.k.moore@sheffield.ac.uk

Abstract

The 2011 release of *Siri* hailed the beginning of a sustained period of impressive advances in the capability and availability of spoken language technology. Subsequent years saw the appearance of competitors such as *Google Now*, swiftly followed by consumer products such as Amazon *Echo*. These devices are seen as the first steps towards more advanced 'conversational' artefacts (especially *robots*). However, evidence suggests that the usage of such voice-enabled devices is surprisingly low, perhaps due to noise in the environment, privacy concerns or manual alternatives.. Another possible contributing factor is that the ubiquitous deployment of *inappropriate* humanlike voices for non-living artefacts might deceive users into overestimating their capabilities, thereby creating a conflict of expectations that ultimately leads to a breakdown in communications. This paper highlights the benefits of providing an *appropriate* voice for a given artefact based on three separate studies. Results are presented that demonstrate the positive impact of a non-human voice and illustrate how 'appropriateness' might be measured objectively. Finally, a worked-example is presented of implementing an appropriate voice for the *MiRo* biomimetic robot. It is concluded that these insights could be important for the design of future generations of voice-enabled artefacts.

Index Terms: appropriate voices, robot voices, speaking artefacts

1. Introduction

After more than 40 years of research into spoken language processing, the 2011 release of *Siri* - Apple's voice-based 'personal assistant' for the iPhone - represented a significant milestone in bringing speech technology to the attention of the general public. It also hailed the beginning of a sustained period of impressive advances in the capabilities of the underlying speech technologies with dramatic improvements in the accuracy of 'automatic speech recognition' (ASR) and the quality of 'text-to-speech synthesis' (TTS). Subsequent years saw the appearance of smartphone-based competitors to *Siri* such as *Google Now* and Microsoft's *Cortana*, swiftly followed by voice-enabled consumer products such as Amazon *Echo* and Google *Home*. These latter devices are seen as the first stepping stones towards more advanced 'conversational' artefacts in the future, in particular 'automonous social agents' (such as robots) - see Fig. 1.

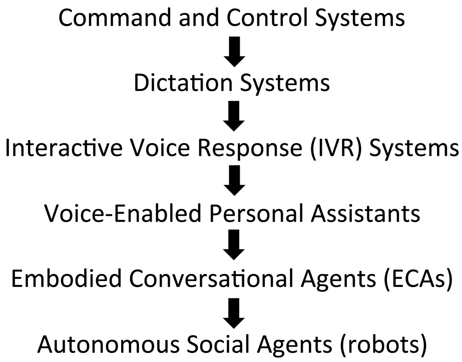

Figure 1: *The evolution of spoken language technology applications from the first 'voice command' systems of the 1970s, through contemporary smartphone-based 'personal assistants' (such as Siri) to future 'autonomous social agents' (i.e. robots).*

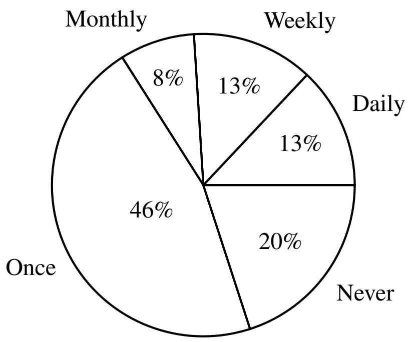

Figure 2: *Speech technology usage on smartphones [1].*

Notwithstanding the popularity of contemporary voice-enabled devices, it appears that actual usage is surprisingly low (see Fig. 2) [1]. Indeed, it seems that voice interfaces maintain their notoriety for "*fostering frustration and failure*" [2].

There are a number of potential explanations for this lack of genuine take-up: e.g. noise in the environment, privacy concerns or manual alternatives. However, it is argued here that another contributing factor could be the ubiquitous deployment of humanlike voices for artefacts that are clearly not human. Not only is this true of mainstream speech-based systems such as *Siri* and *Echo*, but it is also typical to find that robot research laboratories have equipped their devices with off-the-shelf humanlike speech synthesis on the basis that it's "*natural*" that people should wish to interact with a robot using 'normal' speech. The reality is that, when faced with such artefacts, users tend to be deceived into overestimating their capabilities, creating a conflict of expectations that ultimately leads to a breakdown in communications (much like the famous 'uncanny valley' in robotics [3, 4, 5]) - the opposite of what was intended.

In practice, it would be relatively easy to manage users' expectations by giving artefacts an appropriate *non-human*, rather than humanlike, voice. In principle, such an approach would avoid the pitfalls of the 'uncanny valley' by aligning an artefact's visual, vocal and behavioural *affordances* [6, 7, 8], and would create a more 'habitable' interface in line with the ideas expressed in Bruce Balentine's seminal book on the usability of