

9. *Vocal persona*

'Are you talking to me?'

My mother (may she rest in peace!) was in one sense very musical. She could, for example, use a very sad, even miserable, voice to say 'don't worry about me, I'm fine' [◀00]. As a child, I learnt to interpret that sort of statement in two ways. I could either take what she said at verbal face value, i.e. as 'don't worry about me, I really do feel fine'; or else I could privilege the more musical value of her statement, i.e. 'I am feeling sad and miserable'.

The second interpretation was usually nearer the truth than the first, often because of the narrative context of her statement – perhaps after a disagreement with my father (may he also rest in peace!) – and because of her facial expression, body posture and gestures (or, in this case, lack of gesture), all of which aligned with her vocal timbre and intonation but contradicted the meaning of her words.¹ Sometimes it was easier to take her statement at no more than denotative face value, especially if my father, on hearing my mother's suffering tone of voice, were to chide me for not being sensitive enough. I could then quip 'but she told me not to worry' but his reprimand did encourage me to listen more to mother's 'music' and pay less attention to her words. The only trouble with this interpretation was that I would start sizing people up emotionally, on the basis of their 'music' (in the sense of timbre, volume, inflexion, Gestural and facial expression), and I would virtually ignore their words. There were obvious problems with that interpretative strategy too because, if I responded to my mother's plaintive tone, rather than to her words, by asking, in a sympathetic voice, 'what's the matter, mum?', I risked aggravating her irritation and hearing her retort: 'I said I was fine. Why don't you listen to what I say?'

It took me many years to realise that I should actually interpret my mother saying [*plaintive tone* →] 'don't worry about me' [*normal reading* →] as an integral statement, despite its apparent contradictions. It actu-

1. It is difficult to sound miserable with a cheerful face and with lively body movements.

ally meant something like 'I am very sad, poor me, and I find it hard to put on the brave face of self-control I know I should as an adult. So, please pity me, show me some kindness and respect the fact that I at least know I'm supposed to put on a brave face, even if I expect you to see through it'. Yes, I was slow to learn that you could consider the narrative context, the scenario, the body language, the words *and* the music of my mother's complex statements *as a whole* to be grasped instantaneously. It was a musicogenic statement in the same way as the clear musical moods, mentioned in Chapter 2, which were characterised in the form of pallid verbal approximations like DESPERATELY TROUBLED IN THE MIDST OF CALM AND BEAUTY (e.g. Bartók's 4th quartet), or SICK OF THE WORLD AND FEELING ALIVE BECAUSE OF THAT DISGUST (e.g. Nirvana's *Lithium*) or 'DEEPEST SADNESS' AND 'HIGHEST ELATION' (Mozart's 40th symphony).²

The DON'T WORRY ABOUT ME anecdote actualises at least three important issues of musical meaning, the first two of which have been discussed in earlier chapters. This chapter concentrates on the third point.

1. Musical meaning is never created by the sounds on their own. They always exist in a syntactic, semantic and socio-culturally pragmatic context upon which their semiosis depends.
2. Precision of musical meaning does not equal precision of verbal meaning or that of any other symbolic system. Hence, its apparent contradictions of denotation and connotation (see pp. 38-40; 112, ff.) are irrelevant and should be treated as musically coherent.
3. Vocal timbre, pitch, intonation, inflexion, accentuation, diction and volume, plus the speed, metre, rhythm and periodicity of vocal delivery are efficient indicators of the affective disposition communicated by an individual speaker or singer using those means of expression.

The third point is illustrated in clip 00 in which Travis, played by Robert De Niro in *Taxi Driver* (1976), practises his famous question 'ARE YOU TALKING TO ME?' We shall examine this ten-second film extract in detail because it highlights important links between voice and personality.

2. These linguistically contradictory approximations of unequivocal musical mood are explained under 'Synaesthesia and cross-domain representation' on pages 38-40.

Travis, the taxi driver played by De Niro, is appalled by the cruelty and decadence he meets in his job — ‘All the animals come out at night’ is another well-known quote from the film. His character is also that of a nondescript loser who has exercised his second-amendment ‘right’ and acquired a gun to bolster his confidence. Now he feels able to face up to the hoodlums but his low self-esteem still demands that he prepare for confrontation by practising his opening line ARE YOU TALKING TO ME? in the mirror. He utters three variants of that question. Please note that the timings in brackets, below, refer to clip 000 (see Table 6-1, p. 154), not to their position in the actual film.³

It takes De Niro less than two seconds to say the line and he pauses over two seconds between each time he does so (0:30, 0:34, 0:38). Leaving aside gesture, posture and facial expression for the moment and concentrating solely on the sound of De Niro’s voice, minor differences of inflection, intonation, volume and accentuation can be discerned. The first time he asks the question (0:30) his voice is low-key but quite rapid with the quick but substantial rise of pitch normally applied in English to questions expecting the answer ‘yes’ or ‘no’, but it does sound a little bit sudden, as if he had been taken off guard. The second utterance (0:34) is slightly slower, a little more deliberate and has clearer diction, suggesting that the imaginary low-life interlocutor may not have heard him the first time. The third utterance (0:38) of the same question is, once again, quite low-key, but there is this time slightly more emphasis on ‘*me*’ and a little less on ‘*talking*’, this shift in accentuation underlining his personal involvement in the enacted meeting. Apart from these minor variants, it should be noted that De Niro does not raise (the volume of) his voice in anger or frustration and that his is the normal voice of a young, probably white, North American, English-speaking male. In fact, without the narrative context and without De Niro’s use of body language, there is nothing remarkable about his vocal persona in this scene any more than Travis himself is supposed to be a remarkable personality, even though his distinct lack of charisma may be what makes him narratively interesting.

3. <http://uk.youtube.com/watch?v=OL7uc6L5nMQ>

Table 6-1: Clip: Vocal Persona Commutations: ARE YOU TALKING TO ME?

Timing	Visual	Voice
0:00	Logos and titles	robot (repeated)
0:30	TD	<i>original</i> (×3)
0:52	TD, comments	<i>original</i> (×3×3)
1:35	TD	Dad talking to baby (×3)
1:57	TD, comments	Dad talking to baby (×3×3)
2:30	Webcam	<i>Commentator</i> (on baby talk)
3:03	TD, comments	Robot
3:17	TD, comments, close-up	Lecherous (repugnant intimacy)
3:40	TD, comments	Monsters (English, German)
3:57	TD, comments	Pathetic, despondent (×4)
4:05	TD, comments	Posh Southern UK (×4)
4:32	TD, comments	Regular Southern UK (×5)
4:49	TD, comments	Disbelief/ridicule (×5)
5:06	comments	<i>Commentator</i>
5:45	TD	<i>original</i> (×2)
5:51	TD, comments	'It's tacky to be' [or not to be] (×2);
5:56	TD, comments	'Då tar det väl vi.' (×3);
6:02	TD, comments	'Are you watching TV?' (×2)
6:03	TD, comments	'Distracting a bee.' (×2)
6:08	TD, comments	'J'suis bien ton ami' (×2)
6:14	comments	Original ×2;
6:16	comments comments	Surprised/indignant; Big guy, big reverb
6:20	comments	Kiddy robot; Posh Southern UK
6:25	comments comments	chipmunk; exasperated (1); exasperated (2);
6:31	comments	Swedish; disbelief; normal German
6:38	end titles	Estuarian; angry robot; very quiet
6:42	end titles	Posh Southern UK; Liverpudlian
6:46	end titles	despondent; whispered; plaintive
6:50	TD, extreme close-up	lecherous
6:57	TD, inverted colours	slow kiddy robot
6:59	TD, distorted face	evil god
7:02	TD, severely distorted face	monster: 'Sprichst du mit mir?'
7:06	TD, pixelated	robot, repeated
7:15	Webcam	<i>commentator</i> , evil child

Given that this relatively normal, neutral and uncharismatic personality has a correspondingly normal, neutral and uncharismatic vocal persona, it ought to be possible to replace his voice with others in order to understand how certain vocal elements are compatible or incompatible with other simultaneous aspects of non-verbal communication. We'll deal first with the latter in the *Taxi Driver* mirror scene.

The fact that we're in quite a noisy and untidy kitchen and that the young man is white, unshaven and wearing what appears to be an grey flannel or denim air-force jacket tells us quite a lot. It rules out vocal personae who are children (6:20, 6:25, 7:15), women, old men, African-American-sounding or from the higher echelons of society (4:05), unless they're slumming it, of course. It also rules out robots (3:03), death-metal monsters (3:40) or anything else that doesn't look –or sound– like a Caucasian, between 25 and 50 years of age, who is a male member of the popular classes.⁴ But there is much more in the visuals (0:30-0:45) that narrows down the vocal commutation possibilities.

Since De Niro is about one metre away from the camera, alternative voiceovers cannot sound too close (3:17, 3:57) or too far away (6:40). For example, the 'repugnant intimacy' of the lecherous voice (3:17) only works if De Niro's face is in extreme close-up (3:30, 6:50). Obviously, then, one element of vocal persona is its perceived proximity.

The first time (0:30, 0:52) Travis asks the question he is at the far right edge of the screen with his body facing screen left. He turns his head towards us, as if just having heard something coming from the direction of the camera. He looks surprised, his eyebrows are slightly raised and his head tossed back a bit. It is the look of someone literally taken slightly aback. However, there is nothing except the narrative context that rules out the possibility of *pleasant* surprise, which is why the first BABY TALK voiceover (1:35) works well if viewers imagine the camera being the baby's point of view and that the De Niro character is a proud father, surprised and delighted by his infant's happy and communicative gurgling as he walks past.

4. Monster and robot voiceovers work better if you manipulate the visuals (6:50-7:10).

For the second version of the question (0:34, 1:03) De Niro has half turned toward the mirror/camera, tossed his head back a bit more and raised his eyebrows higher. Once again, it is mainly the narrative context that rules out a possibly positive interpretation of Travis's body language and which lead us to believe that this more clearly 'taken aback' posture is more likely to express affront and irritation than surprised delight. Moreover, since there is no smile on his face, since he seems to be 'looking down his nose' at his imagined interlocutor, and since his diction and accentuation are slightly more forceful than before, the BABY TALK voiceover of the delighted dad is less convincing here. Furthermore, the despondent, depressed and weak voiceovers (3:57, 6:46) align badly with De Niro's posture, facial expression, accentuation and diction during these three seconds.

The third version (0:38, 1:25) is gesturally the clearest. His body is turned a little more towards us (the camera/mirror/imagined interlocutor), as he points to his own chest in sync with 'to me'. Yet again, prior knowledge of Travis's character and story will most likely lead viewers to interpret his grin as bemused but insolent, and his hand gesture as expressing personal affront. However, without such prior knowledge and with the addition of a few narrative correctives (baby gurgling, mother going 'aaah!' as if to say 'how cute!'), ARE YOU TALKING TO ME? spoken by a personally delighted and surprised father aligns quite convincingly with this third version of the famous question (c. 1:50, 2:20).

Several vocal persona commutations do not work because of problems with lip sync (synchronisation). For example, stereotypical robot voices (3:03) tend to apply equal durations for each syllable — a non-human speaking trait if ever there was —, while depressed and despondent statements (3:57) are much slower than the rate at which De Niro delivers Travis's famous question in a normal speaking voice. Similarly, whispering and other types of vocal close-up are incompatible not only with the lack of visual close-up in the *Taxi Driver* sequence but also with its speed of delivery: whispering has to be slower than talking because it has to compensate for its lack of voiced consonants and clear vowel sounds, while many types of intimate statement are unsuitable if delivered in a rapid tempo (e.g. 'I love you' at breakneck speed).

+ speed of delivery

