Pre-print of book chapter in *Lowering the Boom: Critical Studies in Film Sound (2008)*, edited by Jay Beck and Anthony Grajeda, pp 36-50. Urbana and Chicago: University of Illinois press

His voice was as intimate
as the rustle of sheets.
Dorothy Parker
(1893-1967)

The proxemics of the mediated voice

In everyday communication a voice is not only a bearer of words and semantic meaning, but is a wonderfully complex instrument to express a range of communicative features. A speaker, for instance, more or less consciously uses his or her voice to signify a communicative 'closeness' or 'distance' towards a listener – something this listener interprets without much effort or even awareness. When a voice is technologically mediated the way a voice sounds is affected, and hence also the way a listener may interpret such aspects as the 'communicative distance' between speaker and listener (both between characters inside a diegetic space and between a mediated speaker and the listening audience).

In analysis of mediated voices, we so far lack terminology to help us distinguish between different spatial aspects of the voice, in order to – for instance – discuss the mentioned expressive and communicative features. This article is thus an attempt to provide some basic tools in this regard. In order to do so, it first must first address a few basic questions: How may one in a fruitful manner analyze the way a voice signals a communicative 'closeness' or 'distance' towards a listener? How does the act of mediation influence the sound of a voice, and the perception of communicative distance? What are the differences between interpersonal (non-mediated) and mediated voices?

Body and voice as signifiers of a communicative relationship

Of all sounds, the voice is the most important in human communication. Research on perception and cognitive processing indicate that recognition and processing of faces and voices appear more specialized than processing of other visual objects and sounds (cf. Spelke and Courtelyou, Eysenck and Keane and also Maasø for an overview of research). As mentioned, the voice is clearly communicating a wealth of information other than mere semantic and referential meaning, such as attitudes, emotions, closeness and distance in relation to other subjects in communication. Such para-linguistic features even seem earlier developed and perhaps more fundamental, than other vocal aspects. Collier, for instance, states that: " . . . the ability to judge emotions through vocal features develops earlier than the ability to judge emotions through facial expressions and body movements and may even be innate" (Collier 141).

The ability of talk to signify a communicative relationship between a speaker and listener rests upon several conditions such as the spatial relationship between speakers, the limited reach of sound in space, and ability of the human voice to vary the address – from an intimate whisper to a distant call. The relationships of bodies and spatial vocal expression, is thus of central importance in communication. Social-semioticians Hodge and Kress even claim that: "Of all the dimensions of the semiotic situation, the most fundamental is the physical relationships of the [bodies of] participants in space" (Hodge & Kress 52).

Spatial listening

Vision is commonly held to be a more precise judge of spatial characteristics in our environment than the audition (cf. Juelez & Hirsh, and Maasø 15-38 for an overview of research). Listening is

most precise concerning horizontal direction, and least accurate when estimating distance. While listening may not accurately distinguish between a speaker placed 6 or 7 feet away – which would be a simple task for vision – we can, of course, easily hear if a voice is spoken 7 inches or 7 feet away, and whether speech is intended for one close by listener or towards a large group.

When evaluating spatial aspects of the voice, three acoustic factors seems of special importance: 1) volume or sound level of the sound 2) frequency characteristic 3) the relationship of direct to reflected sound.

The most important factor for judging the distance of a voice in everyday life is the sound intensity of the voice. Even though our everyday experiences do not tell us that there is a precise inverse relationship between sound intensity and distance (cf. the so called inverse square law in physics), we nevertheless know that sound intensity drop with the increasing of distance. In everyday life we are therefore used to adjusting the sound level of our voice according to whom we address, speaking more softly when addressing a close by listener than a more distanced audience. Likewise, we are all effortlessly taking account of sound intensity when interpreting whether a voice is directed towards us or not. Meeting people talking loudly on a cell phone may thus sometimes lead to a moment of confusion, before we are able to see that they not talking to us, but to an absent listener.

A second important cue for judging the distance of a voice is the frequency characteristics. When a voice is close we hear more of the lower frequencies, because of the so-called equal loudness contours in psychoacoustics (see for instance Handel), while a voice becomes ‘thinner’ and more high-pitched when heard from a distance.

A third important factor in everyday perception of voices is the sound of the room the voice is spoken in. When we are close to a voice, the direct sound will dominate, while

reflections from the walls, ceiling and floor, will dominate with increased distance. The amount of reflection will of course vary with the environment, from extreme amounts of reflection in a large marble surfaced hall to almost none in a snow-covered soundscape.

Earlier analyses of mediated voices

Sound space is commonly discussed in film analysis, and has also at times been among the most commonly debated topics in practical filmmaking (cf. Altman). However, analytical terminology and methodology making meaning of spatial aspects in film dialogue, and mediated voices in general, are largely missing, and hence also specific analyses of aesthetic practices concerning the spatial aspects of mediated voices. For instance, when Altman, Chion, Lastra, Klimek and Truppin use the terms "close up" and "medium close up" sound, none of these indicate how one might distinguish between the two (or other perspectives), or what role technology plays in relation to the spoken voice itself as a sign of spatial relationship between a speaker and listener.[i]

The lack of analytical categories in analyses of sound in film and television is especially striking when considering the wealth of analytical tools, concepts and theories for dealing with visual space and the mise-en-scène of the human body in film and television.

Outside of film and media studies, analyses of vocal expression certainly exist, such as studies of prosody and proxemics (cf. below). Yet, these only describe non-mediated voices, and provide little help for understanding the difference between interpersonal and mediated voices. As a main goal in this article is to better understand the mediated voices and the act of mediation, I will explore the role of mediation in the following, before going into a presentation of the analytical tools suggested to analyze spatial relations in mediated voices.

Mediation and proxemics

In face-to-face communication speaker and listener are always tied together in time and space. The earshot and sound of a voice in space is hence tied to the intensity of a voice, the distance between listener and speaker, the room's acoustical features and the laws of physics regarding the diffusion of the voice in space. Thus, a softly whispered voice has a limited earshot and will rarely have a high ratio of reflected to direct sound. When sound is mediated, the laws of physics and everyday experience are largely suspended. With electro-acoustic sound, the voice may rather be characterized by what R. Murray Schafer has called schizophonia: a split (gr. skhizein) in time and/or space between the production of sound and listening (cf. Schafer 90). The microphone and loudspeaker can thus raise a softly spoken sound to a sound level far outreaching what is possible within non-mediated talking distance. Through the act of mediation sound can thus be broadcasted over vast distances or be recorded for playback in a different time or space. The original loudness of a voice, the sound of the voice and the spatial signature from the room of the speaker thus become individual parameters in mediation; parameters that are all important in evaluating distance and the relationship between body and voice in everyday life are therefore open to technological manipulation and aesthetical choice in mediation.

When anthropologist Edward T. Hall coined the term proxemics in the late 1960s, his neologism sought to emphasize the importance of a spatial proximity and physical relationships between subjects in interaction and communication. Hall showed how distance and social behavior were interrelated, and varied according to whom one communicate with (e.g. a stranger, acquaintance, friend, family, lover) and in what social context (e.g. at work, at home, in bed etc.). Based on extensive studies of social interaction, Hall distinguished between four main proxemic zones, that seemed of special significance when interpreting different kinds of

behavior: an intimate zone (where subjects were within 18 inches of each other), a personal zone (from around 18 inches to 4 feet), a social zone (from 4 - 12 feet) and a public zone (more than 12 feet). Within these four main zones, Hall also discussed a 'near' and 'far' region of each zone. Since my main concern here is not to discuss proxemics as a research area in interpersonal communication, but rather use this as an inspiration for developing useful categories in analysis of mediated voices, I will not go further into Hall's work here. (cf. Hall Proxemics, The hidden dimension, and also Baldassare, for a further account of research on proxemics to follow Hall). For the purpose of this article, I trust that the concept of proxemic zones, 'boundaries' for intimacy etc. makes sense on an intuitive level, and may bring readers to reflect on their own experiences with spatial relations in social situations. I believe most of us, for instance, have experienced some awkward situation when someone trespass our borders of intimacy. Such shared experiences made it possible for the creators of Seinfeld to give this phenomenon a new name – 'close talkers' – with an immediate and broad comic appeal (cf. Seinfeld, episode no. 82, The Raincoats, first broadcast April 28. 1994).

Compared to interpretations of the proxemics of visual representations of bodies in space, interpretation of vocal proxemics to a much larger extent depends on the voice as spoken. Thus, to analyze the proxemics of the voice as a question of microphone placement, and registration of direct and reflected sound, would greatly miss one important goal of this study. It would be comparable with discussing visual framing of bodies filmed in the dark: visual framing it is of little interest for the study of proxemic relationships if one only sees darkness. Hence, if the goal is to understand the proxemic and communicative use of the mediated voice, one need to pay attention to both microphone perspective, the perceived reach or earshot – *and* the use of the voice itself as signifiers of proxemic relationships.

The voice is an incredible flexible instrument, capable of varying the sound level from the barely audible whisper around 25-30 dB SPL at 1 m to a scream up to around 120 db SPL at the same distance (cf. Handel, Holman). This vast dynamic span is more than most acoustic musical instruments can achieve. The voice can therefore quickly vary the intended earshot from an intimate whisper to public call, signaling if an utterance is intended for a close by or distant listener. While the physical body in interpersonal communication is tied to relative slow movements in and out of proxemic zones, the voice may therefore cross several proxemic zones within a split second.

The above-mentioned differences between visual and aural space calls for caution concerning the scope and precision of proxemics as a tool in analysis of mediated voices. It is thus important to bear in mind that the boundaries and categories must be regarded more flexible when analyzing mediated sound, than images. Hall also acknowledged the difficulty in application of aural proxemics (in face-to-face communication), compared to touch and vision (cf. The Hidden Dimension 119). Introducing Hall's analytical apparatus to mediated voices does not make the categories less fuzzy. Thus, the analytical tools presented here should be considered but one of several approaches to the study of the mediated voice. Yet, the lack of other tools and methods nevertheless seems to make the search for some common concepts of spatial relations a necessary step in the study of mediated voices.

Three levels of analysis

Since mediation introduces a schizophonic split between different audible parameters important for evaluating distance in a proxemic sense, this needs to be reflected in categories and analytical

levels applied to the study of the mediated voice. While one in interpersonal communication could do with just the signs and zones of proxemic relations suggested by the voice itself (as Hall does), mediation makes at least two more levels important: that of loudness and what loosely might be considered the 'sound' contributed by the recording technology. I thus, suggest using three levels of analysis, which I suggest calling vocal distance, microphone perspective and intended earshot.

Vocal distance is the term chosen to describe the way the voice 'in itself' signifies a proxemic relationship, as it would be in interpersonal communication. For instance, a whisper is intended for a listener within touching distance.

In mediated talk, the volume fader achieves such an important function in influencing our perception of sound space, that volume is recognized as a separate level of analysis called intended earshot. The term 'earshot' does however not mean the outer limits of an audible sound, as one might usually employ the term. The slightly awkward term 'intended earshot' is instead coined to describe the 'primary' earshot signaled by the volume of the voice, or in other words: When a speaker addresses a large and distant audience, s/he will speak with a loud voice that will also be heard within distances closer to the speaker. The intended earshot in this case will nevertheless be regarded directed at the more distant listener and thus having a public earshot. And even though a whispered voice may be (over)heard at some distance, it is nevertheless clear that it is meant for a close-by listener. The low volume thus is a sign of an intended earshot. Because of the schizophonia introduced with mediation, loudness becomes a parameter possible to manipulate separate from the speakers vocal distance, i.e. the sound of the voice itself. The speaker may even intend the voice for a different proxemic zone than what the production team does. For instance, changing a voice with a vocal distance intended for a close-by listener to an

extended earshot, is often used as a dramatic or comic device in film and television, such as in the scene in M*A*S*H (Robert Altman, USA, 1970) where Margaret Houlihan's intimate pillow talk with Frank Burns is broadcasted over the P.A. system for the whole camp to hear, and at the same time providing Major Houlihan with the nickname 'Hotlips'.

The last analytical level used to describe the proxemics of the voice is called microphone perspective. This level encompasses the mise-en-scène of the voice by the technological apparatus and acoustical characteristics important in judging distance (such as of the direct-to-reflected sound and timbre), with the noted exception of 'volume' or 'intended earshot'. Though the choice of the microphone itself – and the placement of the mike in relation to the speaker – is arguably the most crucial part of the technological factors in mediation of the voice, a range of other technologies may also be important here, such as compressors, equalizers, reverb etc. In order to not make the analytical levels unnecessary complicated, however, these factors are all included under the term microphone perspective, with the possible danger of thus overstating the role of the microphone itself.

Whether a voice is recorded in a closet or a cathedral will affect the sound of a voice in important ways influencing our perception of distance.[ii] To simplify the discussion and tie microphone perspective to a reference which is meaningful for the most of us, the descriptions in figure 1 (cf. the second column) is to be regarded as an ideal type, based on the kind of reflections, reverberation and absorption typical for a living room. Likewise, though microphones may vary in spatial representations of a voice, the description is an attempt to indicate the distance of a non-directive mike, even though the particular voice in question may be recorded by a shotgun mike at 2 meters distance or a radio mike taped to the chin of a

performer. The important point is that increasing the microphone distance will achieve a somewhat thinner sound, while the amount of reflected sound to direct sound will increase.

During my work developing and adjusting the analytical tools presented here, I have simplified Edward T. Hall's categories slightly. As mentioned, Hall used four zones divided into a 'near' and 'far' region. For the two closest regions (the intimate and personal) such a level of detail seems unnecessarily inexpedient dealing with sound, as our ears are poorer judges of spatial details in distance than vision and (especially) touch, where smaller movements make a bigger difference within the most intimate zones. However, I found the distinction of 'far' and 'near' to make sense for both the social and public zones, also since the zones would then correspond more closely with terminology known within visual film analysis. I thus ended with a blend of Hall's zones adapted into the language of image framing: extreme close up (ECU) for the intimate zone, close up (CU) for the personal zone, medium close up (MCU) and medium shot (MS) for the social zone, and long shot (LS) and extreme long shot (ELS) for the public zone, as described closer in figure 1.[iii]

Figure 1 brings short descriptions of each of the three analytical levels proposed. While the description attempts to explain how the vocal characteristics sound, describing the aural qualities of sound in such a way is a near impossible task. The figure is hence also available as a web site bringing concrete examples from film and television picked out to show concretely how I have judged these clips.[iv]

Figure 1: Proxemic zones and analytical levels

PROXEMIC ZONES	**Vocal Distance**	**Microphone Perspective** [v]	**Intended Earshot** [vi]
	SOUND / FUNCTION	SOUND / FUNCTION	SOUND / FUNCTION
INTIMATE (ECU) ≈ 5-45 cm ≈ 2''-18''	Moaning; breath; whispering between lovers; intimate confession; confidentiality; exclusion of other listeners; 'back region' behavior	Direct sound only; non-audible reflections; proximity effect and 'bassy' voices; 'dry' sound; clearly noticeable mouth-sounds (breath, click of tongue etc.)	Normal earshot for whispering; loud speech is intrusive; screams can hurt
PERSONAL (CU) ≈ 45-120 cm ≈18''-4'	Soft 'in-doors' voice; personal conversation between 2-3 friends; shyness; closeness; non-exclusive	Barely audible reflections; a sense of room tone; not particularly 'dry' reflection or 'bassy' speech; audible mouth-sounds and breath, but not 'up front'	Normal earshot for soft speech; whispering is clearly intelligible; loud speech is intrusive; screams can hurt
SOCIAL (near) (MCU) ≈ 1.2-2.2 m ≈ 4'-7'	Soft to regular conversation between 5-6 people; closeness; personal and social mode of address; community; non-intimate	Obvious mix of direct and reflected sound, though direct sound dominates; rich mid-tone sound; some audible mouth-sounds and breath	Normal earshot for soft to regular conversation; whispering is heard, but is not necessarily intelligible; loud screams do not hurt, but are intrusive
SOCIAL (far) (MS) ≈ 2.2-3.7 m ≈ 7' – 12'	From regular to partly raised conversation between several persons, or to a small group; community; extrovert; social; partly public	Marked increase in reflected sound, clear distance and perspective; normal mid-tone sound; barely audible mouth-sounds	Normal earshot in regular to raised conversation; whispers may be heard, but may be difficult to understand; loud screams may be intrusive
PUBLIC (near) (LS) ≈ 3.7-7.6 m ≈12'-25'	Voice raised to many listeners; lecture; public; community; self-confidence; authority; attention	Reflected sound dominates, but not at the expense of intelligibility; slightly 'thin' sound; non-audible mouth-sounds and breath	Normal earshot for raised conversation and public speech; soft speech sound weak; intimate conversation hardly intelligible, but may be heard
PUBLIC (far) (ELS) ≈ 7.6 m < ≈ 26' <	Shout across the street; speech to large audience; scream for help; public; authority; attention; power	Exaggerated reflections; noticeable 'slap' echo may be heard above 12-13 m; thin sound with clearly reduced bass; reduced intelligibility	Normal earshot in very loud speech to the furthest reach of screams; soft speech difficult to understand; whispering barely audible

Figure 1 subtext: the descriptions of the different zones and levels in figure 1 are best understood when compared to examples supplied at the following website: **http://www.media.uio.no/personer/arntm/zones/.**

Snapshots from a study

The following will bring small samples from of a larger study where the analytical terminology presented here was put to the test.[vii] I will mention some general findings related to two of the eight genres analyzed, and go further into an analysis of a short promo from Norwegian TV 2

towards the end of the article in order to high-light a few findings with more general theoretical interest.

Figure 2: Typical proxemic spans found in film dialogue and ads, trailers and promos in TV.

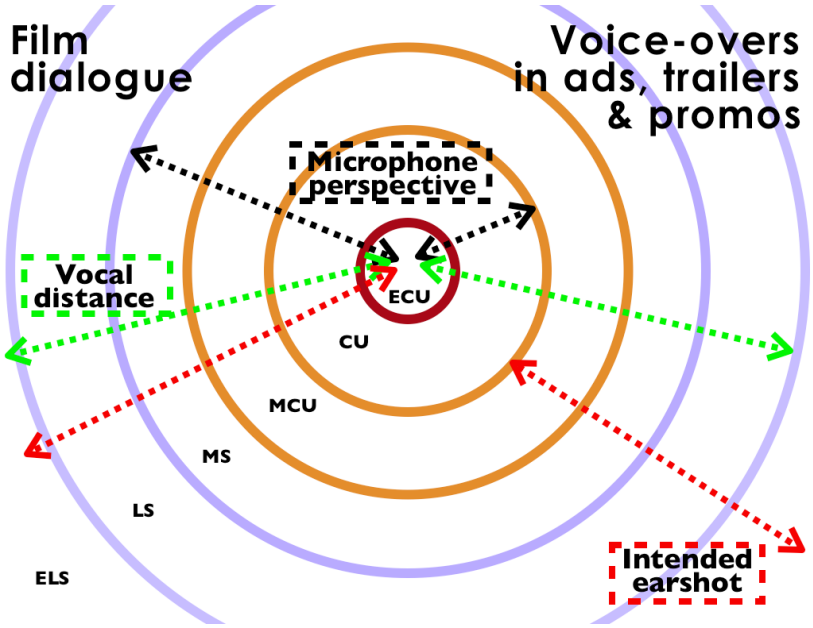

Figure 2 subtext: Figure 2 gives a 'birds view' of the proxemic range of microphone perspective, vocal distance and intended earshot in two of the genres analyzed in Maasø (cf. pp. 149-191 and 269-307). The left hand side of the figure shows distances found in film dialogue, while the right hand side shows the vocal proxemics in voice-overs in ads, trailers and promos. Abbreviations show the six proxemic regions, from ECU (extreme close up) to ELS (extreme long shot).

There is quite a generic variation in the proxemics of the mediated voice as experienced in mainstream film and television. In general, the least variation is found in the analytic level of microphone perspective. Here most of the television genres studied even show a narrower range of mike perspective than in film, with the most common perspective (in the limited material in question here) within the 'personal' proxemic zone (CU). Some genres, such as promos, ads and trailers, as shown above, are very tightly miked with little in any sign of the room surrounding

the voice.[viii] The study thus seems to confirm Altman's general claims for a uniform close-up and medium-close up perspective in film (at least according to the way the categories as defined here). This should come as no surprise since sound recording has become a highly standardized craft (as many have pointed to before), where the perhaps main goal during recording is to come as close as possible with the mike and get a 'clean' recording.[ix] Making the voice seem more distant is simple in post-production or live mixing, by adding reverb and/or lowering the volume.[x] Removing reverb, increasing the gain (without also raising the general noise level) and making the voice sound closer in post-production or live mixing, however, is a radically more difficult task.

The range of intended earshot varies more in both the genres showed in figure 2 as well as others analyzed. Yet the most common earshot across genres in the material analyzed is medium shot, i.e. in the far social zone. Very rarely is the intended earshot intimate or personal. This would also be strange in media concerned with communication. A voice in interpersonal communication is often spoken softly precisely to prevent eavesdropping from other than the intended addressee. Mimicking this is a fairly common narrative device in fiction film and TV-drama (in order to conceal narrative information from characters within the dieresis and from the audience). This is nevertheless rare in film, and virtually unheard of in the majority of broadcast genres.[xi] When Altman uses the term 'for-me-ness' to describe sound perspective in cinema, one might thus ask with Sarah Kozloff: " . . . but who – or what – else would the sound be designed for?" (Kozloff 121).

Vocal distance shows the larges span of proxemic zones of the analytical levels in question. While it is most common to keep speech within the social zone (MCU and MS), the

voice regularly spans over all regions – from intimate whispers to loud shouts. Vocal distance also is the most divisive factor between the genres analyzed. I believe an important interpretative key here is looking at whom the speaker addresses. When one divides addressees into three groups consisting of 1) viewers 2) studio audience or other conversation partners present (whether or not these are found within a fictional diegesis or mediated real life conversation) and 3) both groups 1 and 2 at the same time, the span in vocal distance varies relatively systematic in relation to these three main categories.

For instance, speaker only addressing only the viewers at home (such as program announcers, voice-overs in ads, or certain news announcers) or in the space or diegesis of the recording (such as in interviews, televised debates, or fiction film) usually displays a much narrower span in vocal distance within relative short segments than a talk show host, addressing both a studio audience, viewers at home and guests. Such hosts often display a very dynamic vocal register, where one speaks within a social distance most of the time, but both can lower the voice to ’pull’ a guest into a more intimate mode, or raise the voice to ’play’ the studio audience, before returning to a general social zone again.

Fiction films arguably show the widest variety of vocal modes within a single text, from intimate pillow talk to loud screams of horror, depending on the subject matter and narrative. Nevertheless, many movies, especially mainstream Hollywood movies, display a much more personal vocal register than in most other genres. The vocal distance, and also much more personal than what would be possible in similar surroundings in (non-mediated) interpersonal communication. Much of this is presumably due to great control over the recording situation, and what I also suspect is an underlying ideology (or tradition) of vocal intimacy. Actor Carrie Fisher

has perhaps touched upon this in her humorous description social relations in Hollywood: “You can't find any true closeness in Hollywood, because everybody does the fake closeness so well”.

The schizophonic shout

Nowhere is an intimate address more evident than in advertising and promos in TV. Curiously this is also the genre with the most ‘public’ earshot (cf. figure 1). As more fully discussed in Maasø (pp. 105-148), advertising in television (introduced in Norway in 1988) immediately led to an ‘arms race’ for attention by advertisers. Yet, the loud call for attention – itself an ancient rhetorical trope (attentum parare) – may today be combined with an intimate voice (both concerning vocal distance and microphone perspective), thanks to the schizophonia of mediation. This unprecedented combination of seemingly contradictory proxemic relations is worth a closer look, as I believe it displays some of the unique features of mediation.

Let me use a short promo for Norwegian TV 2 as an example here. The promo features a woman whispering the words ”Se hva som skjer!” (‘Look what’s happening!), while the same words are featured as a colorful graphical representation on screen, the 1.2 seconds it last. When the promo was introduced in 1997 it serving as a ’bumper’ between regular programs and advertisements, and at the same time branding the channel TV 2. Two aspects are of particular interest in this short promo: First of all, it displayed the largest gap between vocal distance and intended earshot ever in Norwegian television, as the whispered voice was also the loudest sound heard on TV until then (cf. an analysis of dB-levels in Maasø). Actually, the sound was so loud that within a few weeks (of furious complaints to the television station) a new version appeared that was 9 dB softer than the original (i.e. the first version was nearly twice as loud as the

second, as experienced by listeners). This brings us to the second interesting aspect of this promo: In the two different versions available, both the vocal distance and microphone perspective are identical, while the intended earshot (volume) differs fairly much. This therefore makes an interesting case to study the interaction between the three analytical levels presented above.

Both these promos (available at http://www.media.uio.no/personer/arntm/tv2/) are both very *intimate* and very *public* at the same time. The promos thus invite us into an intimate communicative relation, and at the same time 'shouts this' out, thus making it possible to hear all the way into the kitchen or bathroom. This intimate shout is only possible through mediation, presenting a form of communication literally unheard of before the advent of mediation. When listening to the two versions mentioned, it becomes clear that this schizophonic split influences the way we perceive the individual aural parameters as well as the overall blend of these: Interestingly, the second (softer) version clearly sounds much more intimate than the first. As only the intended earshot varies in the two versions, this makes for an interesting hypothesis concerning the interaction of the analytical levels in mediated voices.

The schizophonic average

While a listener is able to simultaneously hear both the intimate and distant aspects of the two promos, there is nevertheless also a general appearance in the two – a sort of average or sum of all proxemic levels interacting. Since the mentioned promos obviously vary in 'general' appearance, this has lead me to propose a hypothesis about proxemic relations which I propose calling 'the schizophonic average': The experience of the vocal proxemics of mediated voices is a result of what may be called a schizophonic average of the levels of vocal distance,

microphone perspective and intended earshot. In the media these three levels will both interact in providing a blend of the three and – at the same time – each level plays an independent role. Hence, when there is a gap between the proxemic of the three levels, one can hear both the close and distant aspects simultaneously, while the mix of the levels also provides a proxemic average of the segment or utterance in question. It is because of this average that the second version of the mentioned promo sounds more intimate, since the span between the proxemic regions between the three levels, are less than in the first version of the promo.

Although the comparison of these two promos provides a particularly striking case (given the extreme gap between the three analytical levels) I believe the schizophonic average is a general phenomena working across genres in film and television. This will, of course, have to be put to the test on a wider material than what I have analyzed in Maasø.

Lowering the boom?

The 'schizophonic average' may also shed light on proxemic relations in a wider sense, outside of what is available for analysis. More specifically, it can help us understand historical changes beyond the reach of concrete analysis. For instance, although only a very limited amount of 'old' television broadcasts are archived and available for analysis (at least in Norway), much is known about earlier recording practices. Interviews and document analyses show that there has been a profound change in microphone practices since the early 1980s. Before this time different kinds of microphones on booms were the dominant type of recording strategy in a variety of genres. During the 1980s, however, radio mikes became popular (partly due to increased camera mobility, cranes etc. with increased difficulties avoiding boom shadows in lighting etc.) and is today dominating in most genres, except in one-camera productions in drama and film. With the

diffusion of radio mikes follows a narrower dynamic range than with mikes mounted on booms, and a closer microphone perspective with less reflected sound, since the mikes are very close to the spoken source. Both the use of radio mikes and practice of voice compression also bring out more of the intimate 'mouth sounds', breath etc., that I elsewhere call the back region qualities in the voice (cf. Maasø).

All in all, this has clearly led to an increased intimacy in microphone perspective the last couple of decades in television. Following the logic of the schizophonic average, this would then imply that the proxemics of the voice has become more intimate during the same period.

During the same period, commercial television has been introduced in the Nordic region, as in many other European countries. In the commercial channels ads and promos make out roughly 20% of the total programming hours (compared to more than 1/3 in day time US television). As ads and promos are aurally dominated by voice-overs (cf. Hirdman and Amundsen et al.) and many of these displays a very intimately spoken voice, even the vocal distance has become more intimate in television overall. One might thus say thay television in some ways has raised the boom for good, and yet achieving a more intimate voice than most booms ever did.

Raising one's voice in conversation could mean several things: That a speaker is uncomfortable with the situation and wants to signal further 'distance' to the addressee, that s/he is excited, angry – or simply that the noise level behind the conversation is loud. Describing spatial relations and the proxemics of the voice can thus never take the place of further qualitative analysis. I nevertheless believe it is necessary to develop some terms and categories that reduce the enormous complexity and nuances in the mediated voice in order to facilitate

communication about the voice. Hopefully the vocabulary and categories presented above might thus help open up for further discussion about the role of the voice in communication.

Works cited

Altman, Rick. "Sound Space." Sound Theory / Sound Practice. Ed. Rick Altman. New York & London: Routledge, 1992.

Amundsen, Geir, et al. Den Usynlige Stemmen: Om Bruk Av Voice-over I Reklamefilm. Oslo: Department of media and communication, University of Oslo, 1999.

Baldassare, Mark. "Human Spatial Behavior." Annual Review of Sociology 4 (1978): 29-56.

Bordwell, D., and K. Thompson. Film Art. An Introduction. New York: Mc Graw-Hill Inc, 1990.

Chion, Michel. Audio-Vision. Sound on Screen. New York: Columbia University Press, 1994.
---. The Voice in Cinema. New York: Columbia University Press, 1999.

Collier, Gary. Emotional Expression. Hillsdale, N.J.: L. Erlbaum Associates, 1985.

Eysenck, M.W. , and M.T. Keane. Cognitive Psychology. A Student's Handbook. London: Psychology Press, 1990.

Frith, Simon. "Art Versus Technology: The Strange Case of Popular Music." Media, Culture & Society 8 (1986): 263-79.

Hall, Edward T. The Hidden Dimension. Man's Use of Space in Public and Private. London & Toronto: The Bodley Head Ltd, 1969.

---. "Proxemics." Current Anthropology 9.2-3 (1968): 83-108.

Handel, Stephen. Listening. An Introduction to the Perception of Auditory Events. Cambridge, Massachusetts: MIT Press, 1989.

Hirdman, Anja. Tv-Reklam I Sverige 1990 Och 1995. Maskulinitet, Femininitet Och Etnicitet. Stockholm: Konsumentverket, 1995.

Hodge, Bob, and Gunther Kress. Social Semiotics. Cambridge: Polity Press, 1988.

Holman, Tomlinson. Sound for Film and Television. Boston & Oxford: Focal Press, 1997.

Julesz, Bela, and Ira Hirsh, J. "Visual and Auditory Perception - an Essay of Comparison." Human Communication. A Unified View. Eds. E. E. Jr. David and P. B. Denes. New York: McGraw-Hill, 1972. 283-340.

Klimek, Mary Pat. "Imagining the Sound(S) of Shakespear: Film Sound and Adaptation." Sound Theory / Sound Practice. Ed. Rick Altman. New York & London: Routledge, 1992. 204-16.

Kozloff, Sarah. Overhearing Film Dialogue. Berkeley: University of California Press, 2000.

Lastra, James. "Reading, Writing, and Representing Sound." Sound Theory / Sound Practice. Ed. Rick Altman. London: Routledge, 1992.

Maasø, Arnt. "Se-Hva-Som-Skjer!": En Studie Av Lyd Som Kommunikativt Virkemiddel I Tv. Acta Humaniora Nr. 132. Oslo: Universitetet i Oslo, 2002.

Schafer, R. Murray. The Soundscape: Our Sonic Environment and the Tuning of the World. Rochester, Vermont: Destiny Books, 1994.

Spelke, Elisabeth S., and Alexandra Courtelyou. "Perceptual Aspects of Social Knowing. Looking and Listening in Infancy." Infant Social Cognition. Eds. Michael E Lamb and Lonnie R. Sherrod. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1981. 61-84.

Truppin, Andrea. "And Then There Was Sound: The Films of Andrei Tarkovsky." Sound Theory / Sound Practice. Ed. Rick Altman. New York & London: Routledge, 1992. 235-48.

van Leeuwen, Theo. Speech, Music, Sound. London: Macmillian Press, 1999.

Woram, John M. Sound Recording Handbook. Oslo: Vett & Viten AS, 1995.

---

[i] The focus here will be on the sound of the voice. Although the interaction of vision and audition is of great importance when interpreting proxemic relationships in interpersonal as well as mediated communication (such as discussed by Chion), I will not deal directly with this in this here. The attempt of providing analytical tools to deal with the audible aspects of the mediated voices will thus need to be tested in audiovisual analyses of film and television.

[ii] Manipulating the level of reflected to direct sound, by adding reverb, is another fairly common dramatic device in film and television, for instance in The Third Man (Carol Reed, GB, 1949). In the sequence shortly before Harry Lime (Orson Welles) is shot trying to escape the police in the Vienna sewers, we hear a cacophony of distant voices, all with an extreme amount of reverb, thus also becoming an effective sign of the inner confusion and turmoil that the character is experiencing. Compared to in 'real life' in similar settings, however, the visual distant voices would not have the close up vocal perspective such as many of these voices have, and would also sound 'thinner' than the voices we hear.

[iii] After developing the categories for my research in Maasø (2002), I became aware of Theo van Leeuwen's then newly released (and excellent) book Speech, Music, Sound. Here van Leeuwen interestingly introduces similar categories for what I have called vocal distance (cf. p 24-27), and also with Edward T. Hall's work as inspiration. Van Leeuwen too simplifies Hall's eight regions, and presents five ('intimate', 'personal', 'informal', 'formal' and 'public'). The 'informal' zone in van Leeuwen's terminology, however, entails both what I call medium close up and medium shot (i.e. the near and far region of the social zone). My own view is that it is too reductive collapsing the near and far region of the social zone into one, since precisely this region seems to divide the more informal announcers and more formal news anchor, which in my view is a significant shift. More importantly, though, the main contribution of my own analytical terminology lies not in the adjusting and simplification of Hall's original categories, but in the introduction of the analytical levels of microphone perspective and intended earshot, in order

to deal with the challenge of mediated schizophonia. Though van Leeuwen acknowledges the importance of loudness and microphone perspective, he does not incorporate these into his proxemic categories.

[iv] Listening for the vocal distance may be a good place to start, as the sound varies fairly much within this level. The two other analytical levels require more training in what Chion call's 'reduced listening' (Chion Audio-vision) in order to discriminate between the different zones, especially concerning the category mike perspective. This has, at least, been my own experience, as well as a few students testing out these analytical categories in their own analyses.

[v] The relationship between the sound of the voice and the sound space may vary within the same zone. For instance, a voice with clearly 'mouth sounds', sound of breath etc. would qualify for an ECU mike perspective even though the frequency spectrum is not very bass rich. Similarly, a voice with clear 'proximity effect' (see below), such as in many trailers and commercials, will also be regarded recorded with a ECU mike perspective, although this may not contain any 'mouth sounds', breath or other signs of an intimate mike perspective, since such signs of 'imperfection' is very often edited out of commercials and trailers, as interviews with producers and sound designers in advertising show (cf. Maaso 115 and 164). The proximity effect is a technical term describing an increase in lower frequencies when using a directional mike at short distances. While this has often been regarded as a technical flaw, others have used it as an aesthetical effect to achieve intimacy in radio, popular music and advertising since this type of microphone was introduced in the mid 1930s. Frith argues convincingly that 'crooning' as a poplar music style is intimately connected with this microphone technique, achieving a type of communicative intimacy unheard of previously. (Cf. also Holman 70 and Woram 88-90 for technical and acoustical accounts of the proximity effect).

[vi] Note that any sound within the zones of intended earshot will be possible to hear at closer zones, and will be audible at least two zones outside (with the obvious exception of the far public (ELS) zone.) In other words the boundaries between zones are especially fuzzy within the category of intended earshot. Note also that the examples used in the description of intended earshot point to interpersonal communication. A whispered voice will rarely be audible at distances above 7 meters (≈ 25 feet) in interpersonal communication, while this is quite possible with electroacoustic sound, as the sound clips accompanying the figure clearly exemplifies. Since listening levels will influence the experienced earshot when watching television, it is important to listen at a consistent level when performing analysis. I have chosen to listen at a level where regular news speech is roughly at the same level as under a conversation within a social proxemic zone in interpersonal communication.

[vii] The analysis sampled sound bites of 15-60 seconds in length from different parts of programs in eight genres (news, sports, promo, ads & trailers, fiction film, TV-drama, talk shows, variety shows, portrait interviews) broadcasted on Norwegian TV 2 and NRK 1 one week in March 1997. Slightly more than 200 sound bites were analyzed (cf. Maasø 149-191 and 269-307). The study was not designed to provide findings in a strict statistical sense (which would have demanded a much larger and representative selection), the material is nevertheless sufficiently large and varied to give a sense of the typical *span* of proxemics one finds within these genres.

[viii] It is worth noting that voice-overs in ads, promos and trailers are all recorded in acoustically 'dead' rooms. More importantly, as the mike is per definition not visible in voice-overs, microphones can be placed as close as wanted – which leads to a ECU or CU microphone perspective in most of these recordings.

[ix] This is also one of the most consistent answers in my own interviews with 25 technicians and producers in television, also confirmed by observation of several TV-productions (cf. Maasø).

[x] Adding reverb is also often used as a conventional sign for increased subjectivity or 'inner thought', for instance in many voice-overs in film and television. This also highlights the importance of interpreting vocal sound and distance in relation to the narrative and communicative context.

[xi] Notable exceptions are documentaries and other factual based programming recording interpersonal communication, such as live broadcasts of political hearings and court cases where a witness may lower his/her voice, turn away from the mike and address a lawyer. In many reality series contestants also try concealing conversations from the TV-audience or other contestants present. However, this often fails as radio mikes, booms and concealed mikes, often very well cover them. (cf. Maasø (in progress) Sonic Surveillance in Reality TV).