1. Definition

Lipreading is the process of inferring language patterns using only the visible speech movements of a talker as a source of sensory evidence about the speech. Note that contextual evidence still plays a role and may include not only the physical, social, and linguistic context but also facial, hand, and body gestures of the talker. In fact, highly competent lipreaders need to be exceptionally good users of context.

2. Terminology

a) Visual speech perception - the least confusing - easily distinguished from auditory-visual speech perception.

b) Lipreading - the traditional term - probably introduced because the lips are the most obvious source of information. Potentially confusing because some writers use the term to refer to auditory-visual as well as visual speech perception. Fell into disfavor because it carries the implication that lip movements are the only source of information.

c) Speechreading - an alternative term introduced in the seventies to avoid the implication that the lips are the sole source of information. Potentially confusing because some people think that lipreading and speechreading are two different things.

Take care in your use of terminology. Explain what you mean.

3. Note:

Visual and auditory-visual speech perception include visual, hand, and body gestures to provide part of the contextual evidence.

Auditory and auditory-visual speech perception include loudness, stress, intonation, and voice quality to provide part of the contextual evidence.

All three processes use evidence from the world, social, and linguistic contexts.

3. What is visible?

a) Anatomically - Motorically,

Positions and movements of the lips and jaw are visible.

Positions and movements of the tongue tip and the jaw muscles are partially visible.

Positions and movements of the tongue body, the velum, the vocal folds, the rib cage, and the diaphragm are invisible.
What is visible?
b) Physiologically, vocal fold vibration, nasal resonance, and airflow patterns are not visible.

3. What is visible?
c) Phonetically, Place of articulation of stressed vowels, and of consonants, is mostly visible. Duration and rhythm are partially visible. Place of articulation of unstressed vowels, voicing and manner of articulation of consonants, pitch and intonation, and loudness are all invisible.
3. **What is visible?**

d) **Phonemically**, Stressed vowels are mostly identifiable - but not their duration. Unstressed vowels are not identifiable. Consonants articulated at or near the back of the mouth are invisible (e.g., k, g, h). Other consonants are only identifiable within groups (visemes) that have different places of articulation. Word stress, sentence type, and sentence, phrase, and word boundaries are not visible.

In general, it is estimated that about 1/3 of the important information in speech is visible but justification for the exact figure is difficult to establish.

4. **The viseme**

Lipreading researchers have introduced the term viseme as something analogous to the phoneme. Thus:

a) **Phoneme** - a category of speech sound and movement patterns whose members serve the same purpose in terms of auditorily perceived word meaning. The substitution of a sound pattern from within the same phoneme category cannot change either intended and auditorily-perceived word meaning. The substitution of a sound pattern from a different phoneme category can change both intended and auditorily perceived word meaning.

b) **Viseme** - a category of speech movement patterns whose members serve the same purpose in terms of visually perceived word meaning. The substitution of a movement pattern from within the same viseme category can change intended word meaning but not visually-perceived word meaning. The substitution of a movement pattern from a different viseme category can change both intended and visually perceived word meaning.
5. **Viseme categories** can be described in terms of the constituent phonemes. The exact results, however, depend on the details of experimental procedure and can be talker-specific. Typically, we think of about 5 or 6 consonant visemes:
   i) the bilabial viseme - including the phonemes /p/, /b/, and /m/ (should also include lip closure before or after an utterance).
   ii) The labio-dental viseme - including the phonemes /f/ and /v/
   iii) The lingua-dental viseme - including voiced /dh/ and voiceless /th/
   iv) The protruded lips viseme - including the phonemes /sh/, /zh/, /tch/, and /dge/
   v) The alveolar viseme - including /d/, /t/, /s/, /z/, /l/, /r/, /n
   vi) The invisible viseme - including /h/, /g/, /k/, /ng/.

Note:
   a) /w/ and the vowel /u/ are easily confused, as is /j/ and the vowel /i/.
   b) visemes tend to differ in place of articulation while the members within a viseme differ in terms of voicing, stop-continuance, and nasal resonance.
   c) the addition of a small amount of acoustic information can often resolve the confusion about voicing and sometimes about manner of articulation. It may also aid in the perception of word stress, sentence pattern, and sentence, phrase and word boundaries. Hence the wonderfully synergistic relationship between lipreading and residual hearing.
   d) Tactile exploration of the face, as in TADOMA is another way of resolving confusions about voicing, stop-continuance, and nasal resonance.
   e) Homophenes are two utterances that look alike on the lips but have different meaning - for example bin/pin - choose/shoes - bike/mike. The adjective is “homophenous”.

6. **Written analogy**
   Here is a sentence in which vowels are preserved but each consonant is replaced by a viseme category. See how long it takes you to decode it.
7. Correlates of lipreading competence

Researchers have looked for predictors of speechreading performance. Among postlingually deafened subjects, however, language knowledge can be assumed to be intact. In such populations researchers have failed to find good predictors. Competent lipreaders are not necessarily more intelligent, more extroverted, or better and faster at pencil and paper tasks. Several years ago, a researcher named Shepperd reported strong correlation between the latency of a visually evoked response and a measure of lipreading performance. But the degree of correlation was suspiciously high and further research failed to replicate the finding.

In research at CUNY, we tested the hypothesis that good lipreaders would also be good auditory perceivers of filtered speech. The speech was low-pass filtered at 500 Hz so as to give about the same phoneme recognition scores in CVCs as is found by lipreading. We found no relationship. The best listeners were not necessarily the best lipreaders.

The basic conclusion is that good lipreaders are good lipreaders. Lipreading seems to be a unique task, involving visual spatial-sequential processing of motorically-coded, meaningful linguistic information. There is no other task quite like it.

Correlates of lipreading competence

Among prelingually deafened subjects, spoken language ability and lipreading ability are fairly well correlated - for obvious reasons. Language is what one is trying to perceive. If one doesn’t know language very well, one will not be good at perceiving it.
8. Lipreading instruction

Various "schools" of lipreading instruction sprung up at the beginning of the 20th century. Details will be found in texts on lipreading and aural rehabilitation. The approaches are usually divided into 'analytic' and 'synthetic'. (See, for example, Hull RH (1997), Improving communication for adults through aural rehabilitation: History, theory, and application. In: Hull RH (Ed.). Aural Rehabilitation: serving children and adults (Third Edition) . Chapter 12, pp 177-190. San Diego: Singular Publishing group).

Analytic approaches are those in which instruction focuses on the details of the sounds - learning how they look on the lips and practicing their recognition in isolation and in words (the Mueller-Walle method and the Kinzie approach). The underlying theory is that the lipreader, deprived of much of the acoustic information, must work hard to extract as much information as possible from the visible movements. An underlying assumption is that speech perception is, basically, a bottom-up process in which the first task is to analyze the incoming signal into its basic components before assembling those components into words and sentences. (In practice, the analytic approaches did not stop at the analysis stage but also gave practice at the sentence level).
Synthetic approaches are those in which instruction focuses on creating a meaningful message from the incomplete input signal (the Nitchie approach). Subjects practice understanding complete sentences, paragraphs, newspaper stories, etc. The underlying theory is that the lipreader will never be able to get enough information from the visible movements of speech and must learn how to synthesize complete messages in spite of this. An underlying assumption is that speech perception is, basically, a top-down process in which the first task is to make hypotheses about words and sentences - only examining the incoming signal to the extent needed to accept or reject those hypotheses.

There have also been motorically-based programs in which subjects practice by watching themselves speak in a mirror (The Jena method).

The results of our own work at CUNY have convinced me that the significant variable in lipreading learning is “time-on-task”. To the extent that formal instruction can increase time-on-task, then it is worthwhile. Some subjects may be able to gain all the practice they need in everyday communication situations. Other subjects may require opportunities to practice with simple materials, and lots of support from context, in a risk-free environment.

It is important to remember, however, that no amount of instruction can enable a lipreader to extract, from the visible stimulus, all the information that would be available from the auditory stimulus. But, by ensuring optimum visibility, and paying close attention to the visible stimulus, he may obtain all the information that is there. And by practicing in a comfortable and supportive environment he may learn to accept interpretations with less sensory evidence, to accept a higher probability of error, to accept error without embarrassment, and to increase speed.
9. Is lipreading instruction effective?
There is evidence from the literature that lipreading ability or capacity is bimodally distributed in the population. In other words, there are two kinds of people - actually or potentially competent lipreaders, and mediocre lipreaders. Faced with the challenge, the first group will have, or will quickly acquire effective lipreading skills without formal instruction. The second group, however, will need help to improve their lipreading skills and, even then, will not become competent lipreaders. In other words, instruction may help mediocre lipreaders to become better mediocre lipreaders but will not turn mediocre lipreaders into competent speechreaders.

10. Audio-visual speech perception
Complete deafness is rare and much of aural rehabilitation should include helping deafened individuals realize and exploit the tremendous benefits of combining lipreading with aided hearing. The combination of poor hearing and mediocre lipreading skills can create a competent auditory-visual speech perceiver.

11. The McGurk effect. When presented with conflicting visual and auditory cues (e.g., visual "ga" with auditory "ba") the observer perceives a sort of average "da". The finding has no practical significance because in the real world, auditory and visual cues correspond. It does show, however, that visual and auditory cues are not treated as redundant or conflicting sources of information but are integrated by the perceptual system.
Lessons for pediatric aural rehabilitation

• Hearing and lipreading enhance each other
• The child hears speech movements by listening
• The child sees the same movements by lipreading
• The movements that are harder to hear are easier to see

And vice versa

SO:
- Use hearing and lipreading together for best communication and learning
- And for best lipreading:
  - articulate clearly but naturally
  - make sure the child is watching you
  - remove detractors (e.g. sparkling jewelry)
  - make sure the light is on your face and not in the child’s eyes

Implications for pediatric aural rehabilitation

1. For communication, learning, vocabulary development, a combination of hearing and lipreading is always more effective than hearing alone
2. To optimize lipreading:
   - articulate clearly but naturally
   - make sure the child is looking at you
   - remove detractors (e.g. sparkling jewelry)
   - make sure the light is on your face and not in the child’s eyes
3. For specific activities designed to test, develop, or enhance and listening skills, there may be times when it is appropriate to withhold lipreading.
4. This is best done by sitting side-by-side with the child rather than covering your mouth.
5. There is an extremist “auditory-verbal therapy” in which lipreading is withheld even for learning, communication, and vocabulary development – often by deliberately preventing the child from seeing the talker’s face. This is done in the belief that, if the child learns to lipread he will not develop his hearing skills fully. There is no evidence to support this position. All the evidence suggests that hearing and lipreading enhance each other and do not interfere. As indicated, however, there may be times when it is appropriate to withhold lipreading in order focus the child’s attention on acoustic differences and to pursue specific hearing and listening goals.

Lessons for pediatric aural rehabilitation

cantd.

- Sometimes the focus is specifically on hearing and listening
- At such times it may be appropriate to withhold lipreading
- If so, sit side-by-side rather than cover your mouth