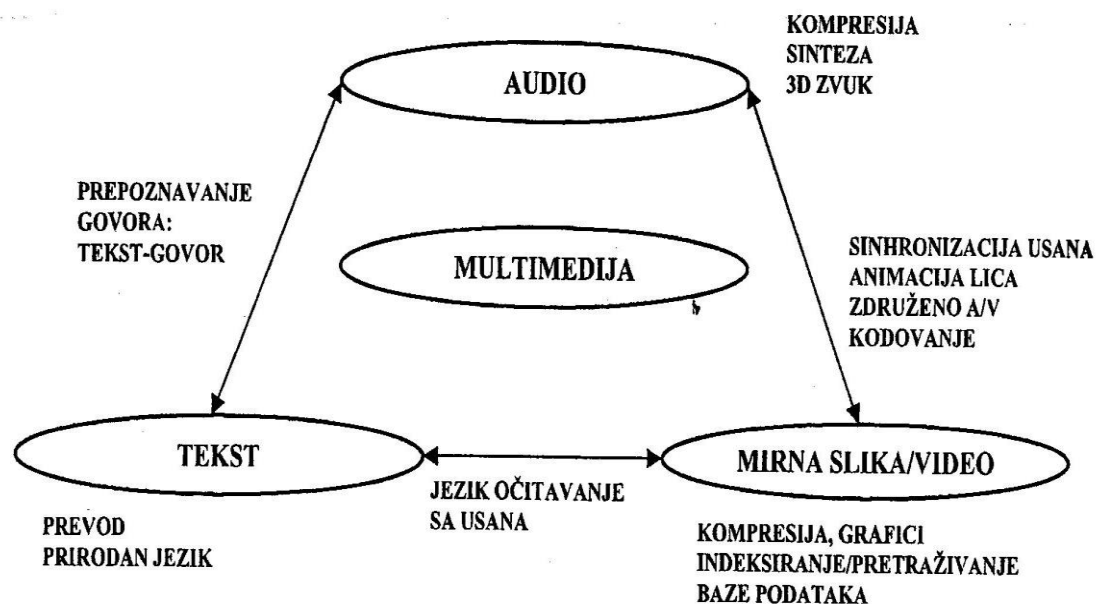


# AUDIO-VIZUELNE INTEGRACIJE

- Multimedijalne komunikacije su objedinjeni tekst, audio, slika i video.
- Audio – vizuelna integracija je od posebnog značaja.
- Koristeći tehnologiju poznavanja govora, analiziraju se govorni talasni oblici kako bi se otkrio izgovoreni tekst. Na osnovu rečenice u tekstu, integriše se glava govornika na osnovu audio–vizuelnog niza i to korišćenjem kompleksne grafike za analizu facijalnog modela kao i sintezu teksta i govora kako bi se proizveo sintetičko akustični govor.
- Dobar primer za audio–video interakciju pri govornoj komunikaciji je očitavanje sa usana koje se naziva još i govorno očitavanje. Koristi se u videotelefoniji i video konferenciji.

- ▶ Cilj prepoznavanja govora je da obezbedi mašinu da bude u stanju da prisluškuje govorne ulaze i individualne reči.
- ▶ Interakcija medija prikazana je na slici



Šlika 4.1.1—Interakcija medija

## Bimodalnost ljudskog govora

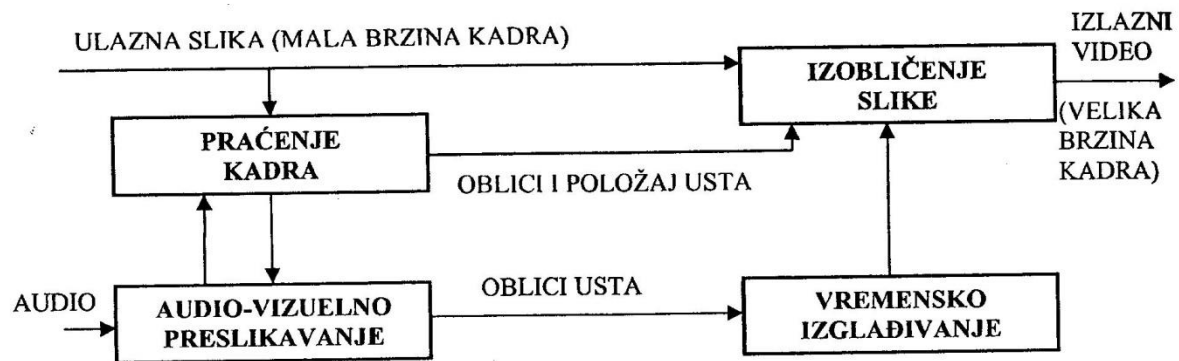
- ▶ Važan je faktor kod projektovanja multimedijalnih komunikacionih sistema (videotelefonija i video konferencije).
- ▶ U slučaju konfliktnih audio i vizuelnih stimulansa, registrovani zvuk ne postoji ni u jednom modalitetu.

Audio+	Vizuelno	Percepirano
BA	GA	DA
PA	GA	TA
MA	GA	NA

- ▶ Produkcija i percepcija govora su po svojoj prirodi bimodalni.
- ▶ Osnovna jedinica u akustičnom sistemu je fonem. U vizuelnom domenu osnovna jedinica (najmanja uočljiva) za meru pokretanja usne je vizem.
- ▶ McGurk–ov efekat znači da govor koji registruje čovek zavisi ne samo od akustičnih znakova, već i od vizuelnih, kao što su npr pokreti usana.
- ▶ Ukoliko rezultati vizuelne percepcije govora utiču na audio govor , radi se o tzv inverznom McGurk–ovom efektu.

# Očitavanje sa usana

- ▶ Očitavanje sa usana zavisi od više faktora:
  - Uslovi gledanja (slabo osvetljenje...)
  - Ugao gledanja
- ▶ Poboljšanje očitavanja sa usana postiže se treningom.
- ▶ Jedno od najvažnijih pitanja u videotelefoniji i videokonferenciji je sinhronizacija usne obzirom na to da je percepcija čovečjeg govora bimodalna. Da bi se rešio ovaj problem možemo izdvojiti informaciju iz govornog signala i obraditi sliku usta tako da se postigne sinhronizacija usne.



Slika 4.3.1—Blok-šema sistema za izdvajanje informacije iz govornog signala i obradu slike usta radi postizanja sinhronizacije usne

## Praćenje usana

- ▶ Za razliku od govornog signala koji je u suštini jednodimenzionalni (1D), vizuelni izlaz je predstavljen 3D video signalom sa dve prostorne dimenzije i jednom vremenskom dimenzijom.
- ▶ Sistem za vizuelnu analizu treba da konvertuje niz slika u značajan parametar.
  - Prva klasa razvrstava sliku usta u jednu od sedam kategorija, na primer u vizeme.
  - Druga klasa meri parametre ili dimenzije na osnovu ulazne slike, na primer, visinu i širinu usana.

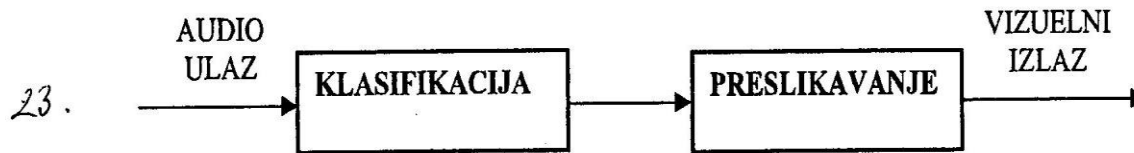


## Audio – vizuelno preslikavanje

- ▶ Vršiti se konverzija govora u parametre koji koriste oblik usta.
- ▶ Kako postoji fizička relacija između oblika vokalnog trakta i proizvedenog zvuka, može da postoji funkcionalna relacija između govornih i vizuelnih parametara.
- ▶ Govor je lingvistički entitet segmentiran u niz fonema. Zatim je svaki fonem preslikan u odgovarajući vizem. Ovakav prikaz pruža najtačniju analizu govora.

# 1. Konverzija na bazi klasifikacije

- ▶ U blok šemi postoje dva dela.
- ▶ U prvom delu, akustični signal se klasifikuje na jednom od mnogobrojnih ulaza.
- ▶ U drugom delu, preslikava se svaka akustična klasa u odgovarajući vizuelni izlaz.



Slika 4.5.1.1—Konverzija na bazi klasifikacije

- ▶ U prvom delu koristi se vektorska kvantizacija da se podele akustički podaci u jedan broj izlaza.
- ▶ Za svaku akustičnu klasu ugrađene su odgovarajuće vizuelne kodne reči kako bi se proizveo vizuelni efekat.
- ▶ Problem je što se ne proizvodi kontinualno preslikavanje već se dobija različiti broj izlaznih nivoa.

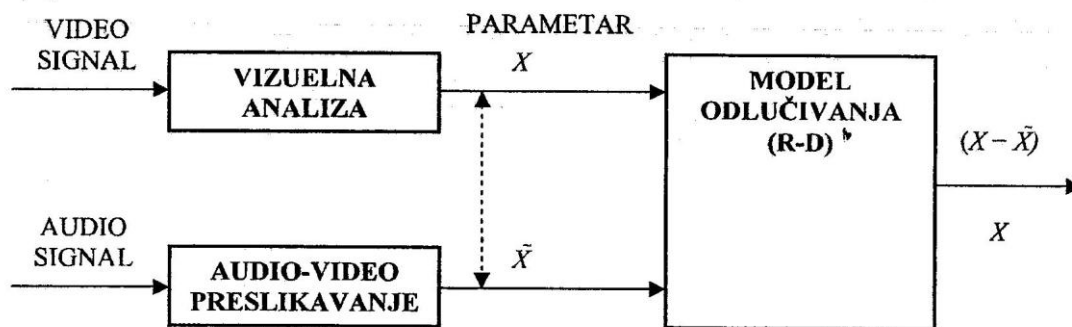
## 2. Audio i vizuelna integracija kod primena za očitavanje sa usana

- Zbog naglog razvoja digitalne video tehnologije, danas je moguće uključiti i vizuelnu informaciju u proces razumevanja govora preko očitavanja sa usne.
- Audio-vizuelni sistem za očitavanje sa usne može se razložiti na tri komponente:
  - Prethodna obrada audio-vizuelne informacije,
  - Strategija prepoznavanja oblika i
  - Integracija strategije.

# Združeno audio–video kodovanje

- ▶ Relacija između audia i videa može se primeniti da se postigne efikasnije kodovanje kako audia tako i videa. Jedan način da se iskoristi ta relacija je prediktivno kodovanje.
- ▶ Ovaj tip kodovanja videa koristi informaciju iz prethodnog video kadra da bi se načinila procena tekućeg kadra. Da bi se na prijemnoj strani obnovio originalni video kadar prenosi se razlika između originalnog i procenjenog signala.

- ▶ Izmereni parametri upoređuju se sa akustičnim parametrima, a model odlučuje o informaciji koja može biti poslata. Odluka se zasniva na kriterijumu funkcije stepena oštećenja informacije.



Slika 4.6.1—Blok-šema unakrsno modalnog prediktivnog kodovanja

# Bimodalna personalna verifikacija

- ▶ Postojeće metode za personalnu verifikaciju zasnivaju se na slici lica ili glasu.
- ▶ Personalna verifikacija samo na bazi mirne slike dovodi do pojave pruga pri kodovanju i varijacija u uslovima raazličitog osvetljaja.
- ▶ Upotreba glasa samo za personalnu verifikaciju nije pouzdana
- ▶ Govorni sistemi mogu da otkažu ukoliko je akustično okruženje u šumovima ili postoji eho.
- ▶ Zato se primenjuju združeno govor i video.